



**HAL**  
open science

## Des bases de données urbaines aux simulations énergétiques - Le projet MApUCE

Serge Faraut, Marion Bonhomme, Nathalie Tornay, Alexandre Amossé, Valéry  
Masson

► **To cite this version:**

Serge Faraut, Marion Bonhomme, Nathalie Tornay, Alexandre Amossé, Valéry Masson. Des bases de données urbaines aux simulations énergétiques - Le projet MApUCE. SCAN'16 (Séminaire de Conception Architecturale Numérique), Sep 2016, Toulouse, France. hal-02157300

**HAL Id: hal-02157300**

**<https://insa-toulouse.hal.science/hal-02157300v1>**

Submitted on 17 Jun 2019

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

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

# Des bases de données urbaines aux simulations énergétiques - Le projet MApUCE

Faraut Serge<sup>(1)</sup>, Bonhomme Marion<sup>(1)</sup>, Tornay Nathalie<sup>(1)</sup>, Amossé Alexandre<sup>(1)</sup>,  
Masson Valéry<sup>(2)</sup>

<sup>(1)</sup>LRA-ENSA Toulouse, <sup>(2)</sup>Météo-France

*RÉSUMÉ : L'énergétique urbaine est un domaine qui couvre de nombreux processus complexes interagissant à toutes les échelles de la ville : le microclimat (l'effet d'îlot de chaleur), la consommation énergétique des bâtiments pour le chauffage ou la climatisation, les comportements énergétiques des habitants, la production d'énergie renouvelable, etc. Peu d'outils tiennent compte de ce niveau de complexité à cette échelle, rendant l'adaptation et l'atténuation aux changements climatiques très difficiles à intégrer dans les politiques urbaines ou les pratiques de conception. Au sein du projet de recherche MApUCE, qui propose une méthodologie pour effectuer des simulations de ces processus d'énergétique urbaine à l'échelle de toutes les communes et villes de France, nous avons développé une approche dont l'objectif est de pouvoir obtenir des données précises sur les caractéristiques et les matériaux des bâtiments. Elle est basée sur la mise en œuvre d'une méthode de détermination automatique des typologies architecturales et urbaines pour toutes les villes de France et d'une base de données architecturale et historique.*

*La méthode de détermination des typologies est une méthode globale et automatisée, dérivée de GENIUS, pour identifier les typologies urbaines à l'échelle du bâtiment et de l'îlot urbain. Elle utilise des indicateurs morphologiques ou sociodémographiques (eux-mêmes déterminés à partir des bases de données nationales) et réalise une classification mettant en œuvre une méthode d'apprentissage supervisé basée sur les « Random Forest ». Ces typologies sont croisées avec les informations de la base de données architecturale, au travers de l'indication de l'usage des bâtiments, leurs dates de construction et leur localisation géographique permettant ainsi de caractériser chacun des bâtiments en terme de matérialité (albédo, système constructif, isolation thermique, etc.) et de systèmes CVC.*

*MOTS-CLÉS : analyse urbaine, typo-morphologie, énergétique du bâtiment, modélisation architecturale, matérialité, évaluation microclimatique, simulation énergétique, classification, adaptation au changement climatique.*

*ABSTRACT : Urban energy is a domain that covers many complex processes that interact at all scales of the city: the microclimate (the heat island effect), building energy consumption for heating or cooling, inhabitants' behavior, renewable energy production, etc. Few tools take account of this level of complexity at this scale, making the adaptation and mitigation to climate change very difficult to integrate into urban policies and design practices. Within the MApUCE research project, which proposes a methodology to carry out simulations of these urban energy processes across all towns and cities in France, we have developed an approach whose objective is to obtain accurate data on buildings' characteristics and materials. It is based on the implementation of an automatic method for determining the architectural and urban typologies for all cities of France and an architectural and historical database.*

*The method for determining the typologies is a comprehensive and automated method to identify urban typologies at the building and the city block scales. The method uses morphological or sociodemographic indicators (determined from national databases) and makes a classification implementing a supervised learning method based on the "Random Forest". These typologies are coupled with the architectural database's information, through the indication of the buildings' usage, their date of construction and geographical location allowing to characterize each of the buildings in terms of materiality (albedo, constructive system thermal insulation, etc.), and HVAC systems.*

*KEYWORDS: typological and morphological urban analysis, building energy, architectural modeling, materiality, microclimate assessment, energy simulation, classification, climate change adaptation.*

## 1. Contexte

L'énergétique urbaine est un domaine qui couvre de nombreux processus complexes interagissant à toutes les échelles de la ville : le microclimat (l'effet d'îlot de chaleur), la consommation énergétique des bâtiments pour le chauffage ou la climatisation, le comportement des habitants, la production d'énergie renouvelable, etc. Peu d'outils tiennent compte de ce niveau de complexité à cette échelle, rendant l'adaptation et l'atténuation aux changements climatiques très difficiles à intégrer dans les politiques urbaines ou les pratiques de conception.

Le projet MAPUCE (Modélisation Appliquée et droit de l'Urbanisme : Climat urbain et Énergie) (Masson 2015) vise à intégrer dans les politiques urbaines et la production des documents juridiques les plus pertinents des données quantitatives ayant trait au microclimat urbain, au climat et à l'énergie. L'objectif principal de ce projet est d'obtenir des données climatiques et énergétiques quantitatives à partir de simulations numériques, en se concentrant sur le microclimat urbain et la consommation d'énergie des bâtiments dans les secteurs résidentiel et tertiaire, qui représente en France 41% de la consommation d'énergie finale. Ces deux aspects sont très liés puisque la consommation d'énergie du bâtiment est très météorologiquement dépendante (comme le chauffage domestique, la climatisation) et que les déperditions ou restitutions thermiques ont un impact notoire sur l'îlot de chaleur urbain (ICU).

MAPUCE propose de développer, en utilisant les bases de données nationales, une méthode générique et automatisée pour générer une classification précise des types de bâtiments et d'îlot à l'échelle urbaine (semblables aux zones climatiques locales « LCZ » (Stewart et Oke 2012)). Pour tous les îlots localisés en France, cette classification permettra de définir des paramètres architecturaux, urbains, géographiques et sociologiques nécessaires pour les simulations énergétique et de microclimat, en particulier l'outil de simulation énergétique « Town Energy Balance »-TEB (Masson 2010) couplé à son « Building Energy Module »-BEM (Bueno 2012) pour MAPUCE. Cet article présente plus spécifiquement la méthode mise en œuvre pour la détermination des caractéristiques architecturales et de la matérialité des bâtiments.

## 2. L'approche

Les simulations énergétiques du bâtiment et urbaines nécessitent de connaître au mieux les caractéristiques des éléments qui la composent, aussi bien leur géométrie que les matériaux de leur enveloppe ou les propriétés de leur environnement.

Nous proposons de développer une méthode de production d'une représentation simplifiée de la ville constituée de l'assemblage d'archétypes d'îlots urbains. L'automatisation de la méthode et le caractère générique des archétypes permettent la représentation d'un large panel de villes françaises. Cette méthode se décompose en plusieurs traitements distincts dont: l'extraction de données à partir de bases de données nationales, une chaîne de production d'indicateurs, une analyse architecturale des bâtiments typiques et une analyse automatique des îlots. Les données sont extraites à partir de bases de données nationales de différentes natures et origines. Afin d'obtenir une méthodologie applicable pour toutes les villes de France, nous avons utilisé seulement les bases de données ayant une couverture globale sur tout le territoire et possédant une certaine homogénéité. Cela inclut les bases de données géo-spatiales BD-TOPO (de l'IGN), celles du parcellaire mais aussi des bases de données proposant des informations sociodémographiques (de l'INSEE) plus liées aux usagers eux-mêmes, à la qualité ou aux conditions d'usages des bâtiments.

Cependant ces différentes données sont obtenues sous des résolutions spatiales distinctes. La résolution spatiale la plus fine est celle du bâtiment, avec quelques informations géométriques ou morphologiques (en particulier l'emprise au sol ainsi que la hauteur du bâtiment pour la base BD-TOPO). Les informations sociodémographiques sur les habitants (comme l'âge des habitants, la composition du foyer) ou sur l'usage du bâtiment (comme le type d'installation de chauffage ou la date de construc-

tion) sont de nature statistique ou seulement précises à une échelle basée sur des zones de population comme l'IRIS (zones correspondant à une population approximative de 2000 habitants).

Afin de concilier ces différentes résolutions, l'unité spatiale de référence qui a été choisie est celle de l'îlot urbain, naturellement compréhensible et adaptée aux urbanistes, tout en étant très proche de la structure urbaine de la ville. Les îlots urbains sont obtenus par agrégations de parcelles contiguës sans être contraints *a priori* par les limites de ses voies de circulation et de ses frontières physiques, à l'aide d'une tessellation de Voronoï (Plumejeaud 2015).

Pour chaque îlot urbain, des indicateurs de nature sociodémographiques ou typo-morphologiques sont ensuite calculés. Quand cela est nécessaire, les données extraites disponibles à des échelles supérieures sont désagrégées à l'échelle de l'îlot urbain, selon des critères appropriés (par exemple de surface de plancher). Le calcul des indicateurs typo-morphologiques définissant les caractéristiques géométriques et morphologiques des bâtiments et du tissu urbain est réalisé à l'aide de requêtes spatiales codées sous OrbisGIS/H2GIS (Boscher 2015), aux trois échelles spatiales dès que cela est approprié (le bâtiment, le bloc (bâtiments contigus) et l'îlot). Au total, 84 indicateurs (dont le temps de traitement est raisonnable à l'échelle urbaine) ont été calculés et ensuite regroupés sous la forme d'une base de donnée relationnelle géo-spatiale.

Parallèlement à ces calculs d'indicateurs, une analyse architecturale des bâtiments typiques de France a été réalisée. Elle a tout d'abord permis de définir 10 archétypes architecturaux représentatifs des bâtiments français. Par la suite, elle a permis d'obtenir une « base de données architecturales » sur la matérialité de ces archétypes (Tornay 2015). L'identification automatique de ces archétypes est basée sur la méthode GENIUS (GENérateur d'Ilots Urbains) (Bonhomme 2013). Le principe est d'utiliser les indicateurs morphologiques déterminés à partir des bases de données nationales pour réaliser une classification, à l'échelle du bâtiment et de l'îlot urbain, mettant en œuvre une méthode d'apprentissage supervisé basée sur les « Random Forest ».

La mise en relation, d'une part des typologies architecturales et urbaine, d'autre part des informations connues extraites de la base de données géo-spatiale, sera utilisée pour fournir les données nécessaires aux outils de simulation énergétique à l'échelle urbaine.

### 3. Base de données architecturales

La base de données architecturales est le résultat de l'association d'une étude sur les typologies des formes urbaines pour les villes françaises avec une étude sur la matérialité des systèmes constructifs.

La première étape de cette approche a été de prédéfinir ces typologies en se basant sur les typologies initialement créées dans GENIUS, et en l'affinant par une étude bibliographique de la littérature technique et les résultats d'une enquête auprès des professionnels chargés de l'aménagement de l'urbanisme d'agglomérations urbaines en France.

Une dizaine de typologies représentatives ont été identifiées (voir tableaux de la Fig. 1).

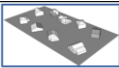
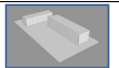
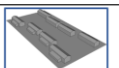

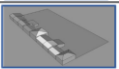

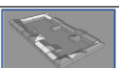


1	PAVILLON DISCONTINU PD		La typologie "pavillon discontinu" correspond aux îlots composés de bâtiments d'au moins quatre façades, RdC ou R+1, souvent implantés au centre de chaque parcelle.	6	IMMEUBLE CONTINU SUR ILOT OUVERT ICIO		La typologie "immeuble continu sur îlot ouvert" correspond à un ensemble de bâtiments en partie alignés sur rue.
2	PAVILLON SEMI-CONTINU PSD		La typologie "pavillon semi-continu" correspond aux bâtiments de type maison jumelée, implantés dans des lotissements ou cités-jardin.	7	IMMEUBLE CONTINU SUR ILOT FERMÉ ICIF		La typologie "immeuble continu sur îlot fermé" correspond aux bâtiments dont l'implantation constitue un îlot fermé.
3	PAVILLON CONTINU SUR ILOT OUVERT PCIO		La typologie "pavillon continu îlot ouvert" correspond au bâtiment de type maison en bande, ou maison de ville, mitoyenne sur deux faces, en alignement sur rue.	8	BATIMENT DE GRANDE HAUTEUR BGH		La typologie "immeuble de grande hauteur" correspond aux bâtiments d'au minimum 12 niveaux appelés aussi « tour », « barre d'immeuble » ou « gratte-ciel ».
4	PAVILLON CONTINU SUR ILOT FERMÉ PCIF		La typologie "pavillon continu îlot fermé" correspond aux bâtiments à patio en bande, constructions en châtresse dans les centres urbains, à l'habitat intermédiaire.	9	BATIMENT D'ACTIVITE BA		La typologie "bâtiment d'activité" est représentée par les bâtiments industriels, commerciaux ou agricoles voire les équipements sportifs.
5	IMMEUBLE DISCONTINU ID		La typologie "immeuble discontinu" correspond aux bâtiments généralement implantés au centre de l'îlot avec quatre façades.	10	ILOT INFORMEL	—	La typologie « îlot informel » correspond aux constructions éphémères, non répertoriées par les cadastres.

Fig. 1: tableau des typologies architecturales et urbaines représentatives

La deuxième étape a consisté à décliner les matérialités possibles pour chacun de ces archétypes en fonction de : (1) la localisation géographique, (2) la période de construction et (3) l'usage du bâtiment (bureaux ou logements par exemple). Ce travail prend la forme d'une base de données relationnelle paramétrable définie au travers de 4 variables d'entrée (en ajoutant la typologie). La base de données permet ainsi de caractériser chacun des systèmes constructifs (murs, planchers et toits) et des matérialités correspondants rendant possible des simulations d'énergie et de microclimat à l'échelle urbaine ( voir Fig. 2).

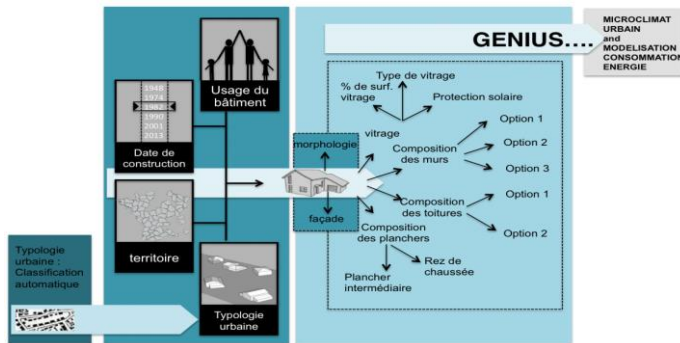


Fig. 2 : intégration des données à l'échelle du bâtiment (source Tornay 2015)

#### 4. Analyse et classification typo-morphologique

Nous avons privilégié pour l'analyse typo-morphologique automatique une approche basée sur une classification supervisée. En effet, les classes de typologie urbaine à reconnaître sont déjà connues et il s'agit alors d'utiliser les indicateurs disponibles pour catégoriser chacun des « individus » des diverses échelles (bâtiments et blocs de bâtiments) dans la classe correspondante.

La classification supervisée se décompose en deux phases : une phase d'apprentissage puis une phase de prédiction. La phase d'apprentissage consiste à créer des procédures de classification à partir d'individus dont on connaît à la fois les variables et le classement. Une fois ces règles établies, la méthode de classification peut alors classer un nouvel individu à partir des seules variables qui lui sont propres.

La phase d'acquisition des données typologiques d'apprentissage est cruciale puisque dépend d'elle la qualité du résultat de la classification finale (éviter le risque de sur- ou sous-échantillonnage ou permettre une validation entre observations et prédictions).

La consistance des échantillons doit aussi être assurée. Pour cela un protocole d'identification typologique de terrain (voir Fig. 3) a été défini pour éviter les mauvaises interprétations possibles sur les observations de terrain, essentiellement menées au travers d'images photographiques aériennes, panoramiques ou 3D (Google Map, StreetView, Google Earth, Géoportail). Ce protocole d'identification a été appliqué pour un ensemble de 7 zones d'études choisies pour leur diversité et leur représentativité.

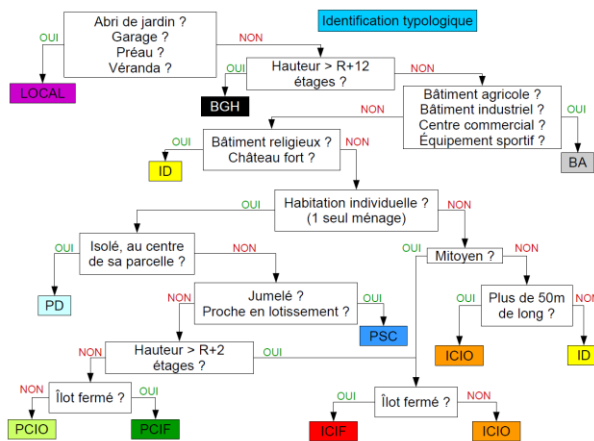


Fig. 3 : Protocole d'identification typologique d'un bâtiment

Dans notre approche nous avons choisi de réaliser l'apprentissage à l'échelle du bâtiment, en ayant comme objectif intermédiaire d'affecter une typologie à cette échelle et d'en déterminer ensuite une typologie à l'échelle de l'îlot. En effet, en raison de leur définition, il est fréquent d'obtenir des îlots hétérogènes contenant des types de bâtiments différents. Nous proposons de déterminer la proportion de chacune des typologies dans l'îlot (en ramenant à leur surface respective par rapport à la surface globale) et d'affecter à l'îlot le type majoritaire (simplification admise pour les calculs d'énergétique urbaine).

Afin de ne pas privilégier une méthode de classification particulière, il a été évalué les résultats de 6 méthodes de classification différentes : Arbre de décision (CART, Classification And Regression Tree), Forêt aléatoire (Random Forest), Analyse factorielle discriminante (AFD) ou analyse discriminante linéaire (LDA, Linear Discriminant Analysis), l'Analyse discriminante quadratique (QDA, Quadratic Discriminant Analysis), K plus proches voisins (KNN, K Nearest Neighbor), Machine à vecteurs supports (SVM, Support Vector Machine).

Un jeu spécifique de données (données atypiques) a été constitué pour tenir compte de certains éléments comme la très grande hétérogénéité des îlots, des erreurs visibles contenues dans les données spatiales issues de la BD-TOPO (fusion de bâtiments différents en une seule entité) en essayant autant que possible d'avoir autant de bâtiments dans chacun des types, en rajoutant si besoin des bâtiments normaux (typiques).

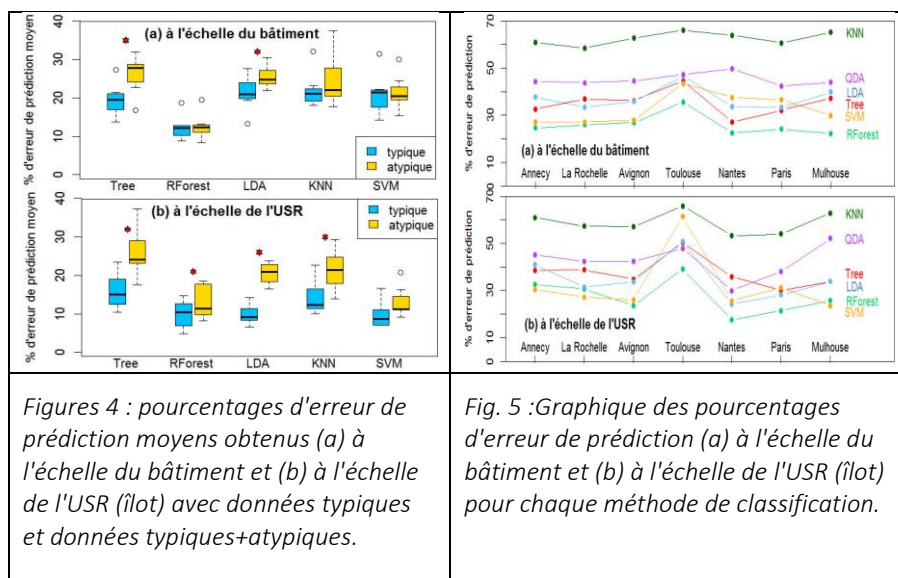
De même, un onzième type « local annexe » a été ajouté pour tenir compte des locaux annexes (garage, abri de jardin, etc.), présents dans les cadastres et devant être attribués à un type.

Pour estimer la qualité du modèle prédictif, nous avons implémenté une validation 70/30. Cette validation consiste à prendre au hasard 70 % des données de chaque type pour réaliser l'apprentissage de la classification, puis à prédire les 30 % des données de chaque type restant avec le modèle de classification créé. Les pourcentages d'erreur de classification des bâtiments et des typologies majoritaires des îlots sont ensuite calculés de manière statistique en croisant les types prédits et observés tout en comparant les erreurs de classification entre données typiques et les données atypiques (voir figures 4).

Afin de déterminer le meilleur modèle global pour toutes les zones d'études, des combinaisons de zones d'apprentissage et de zones de vérifications ont été réalisées (validation sur données exogènes). Pour chacune des méthodes de prédiction, pour chacune des sept zones d'étude, nous prédisons la typologie des données atypiques en utilisant les données atypiques de six autres zones d'études pour la phase d'apprentissage. Les minimums d'erreur de classification observés sont de 22,12 % (pour Mulhouse) à l'échelle du bâtiment et de 17,52 % (pour Nantes) à l'échelle de l'îlot (voir Fig. 5). Après analyse de ces résultats, nous avons choisi de conserver la totalité des sept cas d'étude pour essayer de diminuer le pourcentage d'erreur de prédiction à l'échelle du bâtiment à moins de 10 % avec la méthode de classification RandomForest, et d'optimiser son résultat en influant sur les para-

mètres modifiables (Brostaux 2005), en particulier la suppression des variables non pertinentes, le nombre de variables présélectionnées, le nombre d'arbres constituant la forêt. Cette optimisation du modèle a permis de faire baisser le nombre de variables prédictives de 84 à 74 arbres, de vérifier que le nombre optimal d'arbres de la forêt est de 500, et de fixer à 7 le nombre de variables présélectionnées tirées aléatoirement à chaque nœud de chaque arbre.

Ce modèle optimisé possède alors un pourcentage d'erreur de seulement 11,06 % à l'échelle du bâtiment et de 12,52 % à l'échelle de l'USR d'après une moyenne sur 100 validations 70/30. On est ainsi capable de prédire la typologie urbaine d'un bâtiment en France à 88,94 %.



## 5. Discussion et conclusion

Cette approche confirme que la méthode de classification et d'identification par forêt aléatoire donne de bons résultats. Nous avons également analysé les sources d'erreurs possibles (par exemple en observant la matrice de confusion entre le type de bâtiment observé par type de bâtiment prédit) pour évaluer leur impact possible sur la précision en terme de simulation de performances énergétiques.

Elle permet également d'améliorer les résultats du modèle de prédiction en laissant la possibilité de rajouter facilement de nouveaux indicateurs plus pertinents ou plus évolués.

Pour conclure, dans ce travail nous avons proposé une méthode pour produire des données urbaines de manière automatique et sur tout le territoire français. Les simulations qui en découleront dans le cadre du projet MApUCE vont permettre de porter un tout autre regard sur l'énergétique urbaine : cibler des zones vulnérables au réchauffement climatique, prioriser les quartiers nécessitant une réhabilitation énergétique, identifier des aménagements urbains exemplaires, etc.

**Remerciements :** Ce travail bénéficie d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-13-VBDU-0004.

## 6. Bibliographie

- Bonhomme, M. (2013), *Contribution à la génération de bases de données multi-scalaires et évolutives pour une approche pluridisciplinaire de l'énergétique*, thèse, Université de Toulouse, France.
- Bocher Erwan, Petit Gwendall, Fortin Nicolas, Palominos Sylvain (2015), *H2GIS a spatial database to feed urban climate issues*. 9th International Conference on Urban Climate (ICUC9), Jul 2015, Toulouse, France. 2015.
- Brostaux, Y. (2005). *Etude du classement par forêts aléatoires d'échantillons perturbés à forte structure d'interaction*, Mémoire de thèse.
- Bueno, B., G.Pigeon, L. Norford, K. Zibouche, C. Marchadier (2012), *Development and evaluation of a building energy model integrated in the TEB scheme*, Geosci. Model Dev., 5, 433–448.
- Cutler, D. R., Edwards, J. T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., Lawler, J. J., (2007), *Random Forests for Classification in Ecology*. *Ecology* 88(11) : 2783–2792.
- Hecht, R., Herold, H., Meinel, G., Buchroithner, M., (2013), *Automatic derivation of urban structure types from topographic maps by means of image analysis and machine learning*. In: Buchroithner, M. et al. (Eds.): 26th International Cartographic Conference.
- Masson V. (2000), *A Physically-based scheme for the Urban Energy Budget in atmospheric models*. *Boundary-Layer, Meteorol.*, 94, 357-397
- Masson V., C. Marchadier, L. Adolphe, R. Aguejdad, P. Avner, M. Bonhomme et al. (2014), *Adapting cities to climate change : a systemic modelling approach*. *Urban Climate*, 10, 407-429, doi:10.1016/j.uclim.2014.03.00
- Masson Valéry, Hidalgo Julia, Amossé Alexandre, Belaid Fateh, Bocher Erwan, et al.. (2015), *Urban Climate, Human behavior & Energy consumption: from LCZ mapping to simulation and urban planning (the MapUCE project)*, 9th International Conference on Urban Climate, Jul 2015, Toulouse, France. Proceeding of the ICUC9 - 2015
- Plumejeaud-Perreau Christine, Poitevin Cyril, Pignon-Mussaud Cécilia, Long Nathalie (2015), *Building Local Climate Zones by using socio-economic and topographic vectorial databases*, 9th International Conference on Urban Climate, Jul 2015, Toulouse, France, 2015.
- Stewart, I. D., et T. R. Oke. (2012), *Local Climate Zones for Urban Temperature Studies*, *Bulletin of the American Meteorological Society* 93 (12): 1879-1900. doi:10.1175/BAMS-D-11-00019.1.
- Tornay N., Bonhomme M., Faraut S., (2015) *GENIUS, a methodology to integrate building scale data into urban microclimate and energy consumption modeling*, ICUC 9, Toulouse, juin 2015.