



HAL
open science

Mixture kriging for granular data: The case of energy performance certificate prediction

Marc Grossouvre, Didier Rullière, Jonathan Villot

► To cite this version:

Marc Grossouvre, Didier Rullière, Jonathan Villot. Mixture kriging for granular data: The case of energy performance certificate prediction: The article is published in Science Talks as a video.. Science Talks, 2023, pp.100279. 10.1016/j.sctalk.2023.100279 . hal-04265703

HAL Id: hal-04265703

<https://hal.science/hal-04265703>

Submitted on 23 Feb 2024

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.



Mixture kriging for granular data: The case of energy performance certificate prediction

Marc Grossouvre^{a,*}, Didier Rullière^b, Jonathan Villot^c

^a Mines Saint-Etienne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, Département GMI, Espace Fauriel, 29 rue Ponchardier, F - 42023 Saint-Etienne, France

^b Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, Département GMI, Espace Fauriel, 29 rue Ponchardier, F - 42023 Saint-Etienne, France

^c Mines Saint-Etienne, Univ Lyon, CNRS, Univ Jean Monnet, Univ Lumière Lyon 2, Univ Lyon 3 Jean Moulin, ENS Lyon, ENTPE, ENSA Lyon, UMR 5600 EVS, Institut Henri Fayol, F - 42023 Saint-Etienne, France

ARTICLE INFO

Keywords:

Climate
Energy
Multi-scale
Area-to-point
Kriging
Areal data

ABSTRACT

Energy transition planning requires an in-depth knowledge about a the building stock. To achieve this, data is collected from multiple sources, at multiple scales, with constraints such as privacy policies. This results in data being available for given areas of space such as a land plot. This is the case of Energy Performance Certificate (EPC) database which is anonymized and geo-localized with postal address. EPC prediction has been studied up to now mostly by heat engineers resulting in an effort to classify buildings into typologies. These approaches give good results at a statistical scale (not less than neighbourhood) but poor results at building scale. This paper presents a way to treat EPC as a geo-localized information.

In this paper, we first show that it is possible to apply Kriging methodology to random fields observed at random locations. This new model is referred to as Mixture Kriging. While the usual Gaussian setting is lost, conditional mean, variance and covariance can still be derived. This new model gives interesting results in EPC prediction at building level which is a prerequisite to target renovation efforts. The specific case of a French city is taken as an example.

Video and Presentation to this article can be found online at <https://doi.org/10.1016/j.sctalk.2023.100279>.

* Corresponding author at: U.R.B.S. SAS, Bâtiment des Hautes Technologie, 20 Rue Professeur Benoît LAURAS, 42000 Saint-Etienne, France.
E-mail addresses: marc.grossouvre@emse.fr (M. Grossouvre), didier.rulliere@emse.fr (D. Rullière), jonathan.villot@emse.fr (J. Villot).

Figures and tables

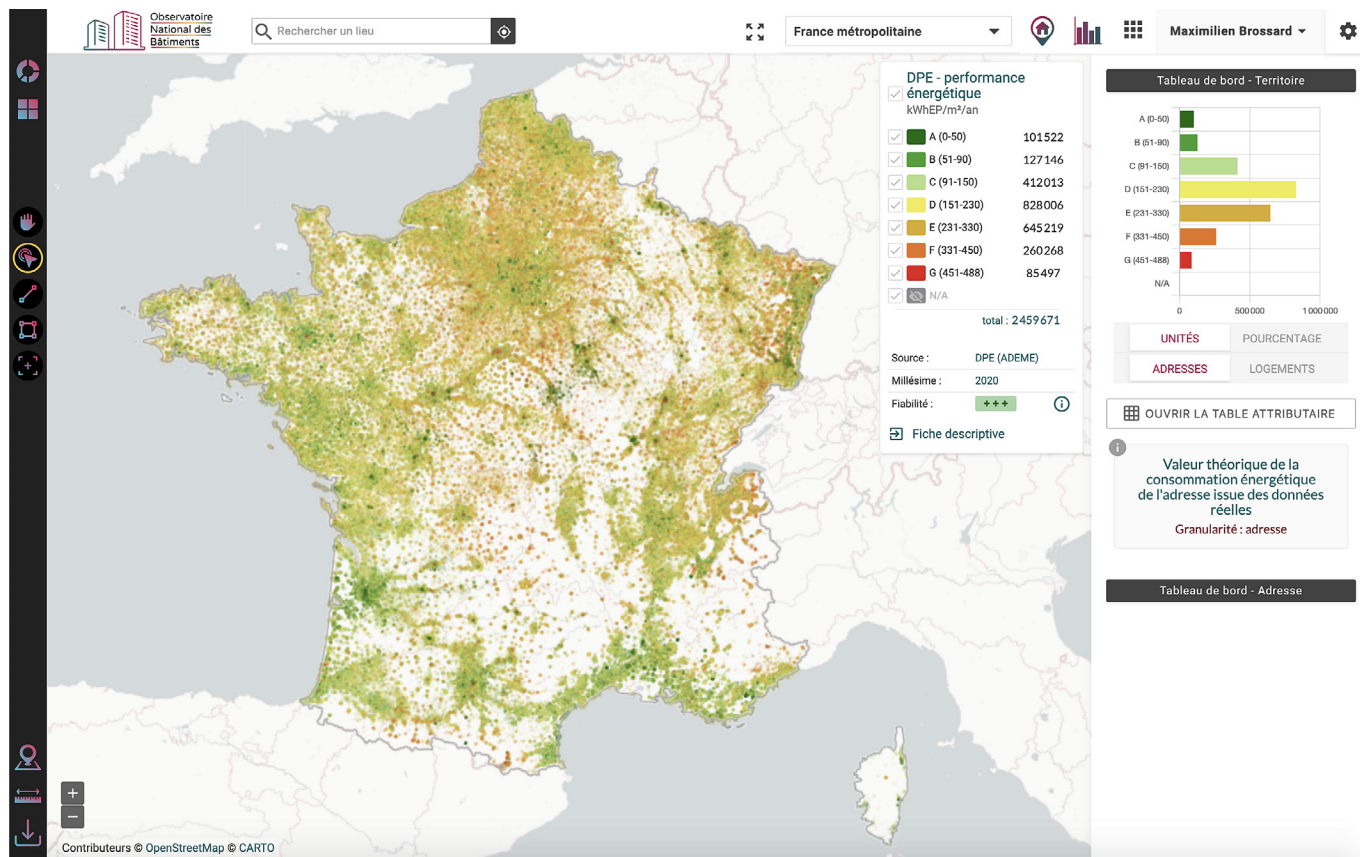


Fig. 1. Screen capture of the National Building Observatory (ONB) giving the distribution of buildings' energy performances in France.

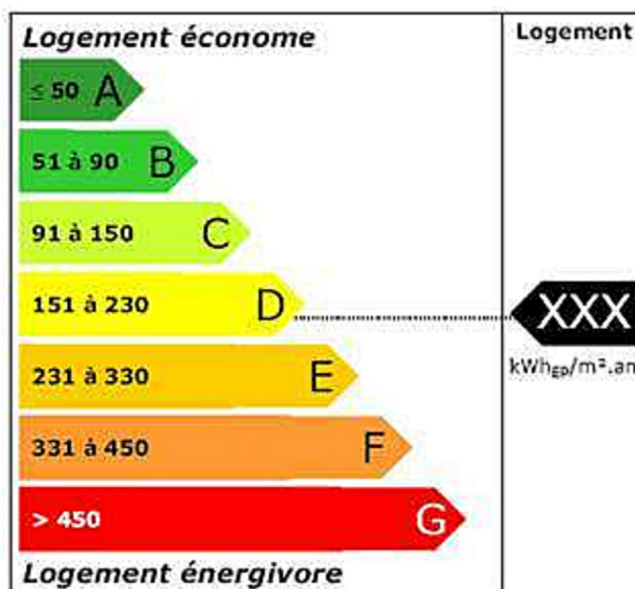


Fig. 2. French EPC vignette.

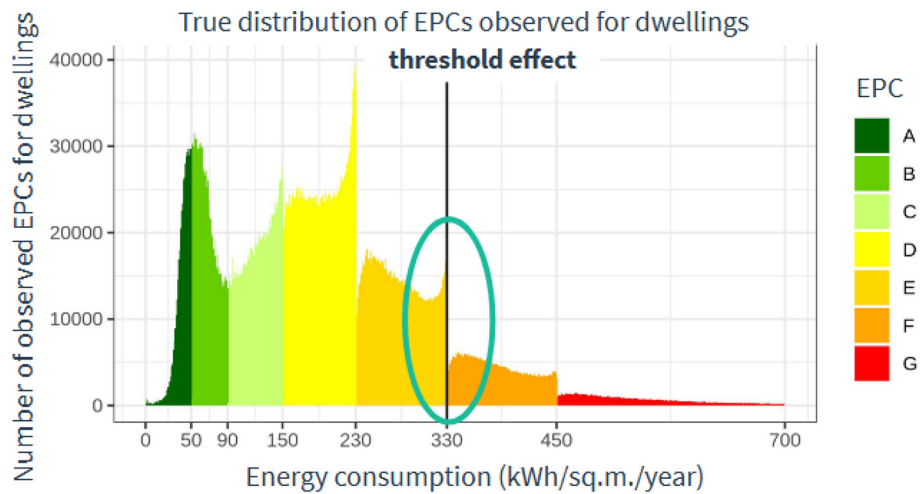


Fig. 3. Histogram of energy consumptions computed to estimate the EPC label.

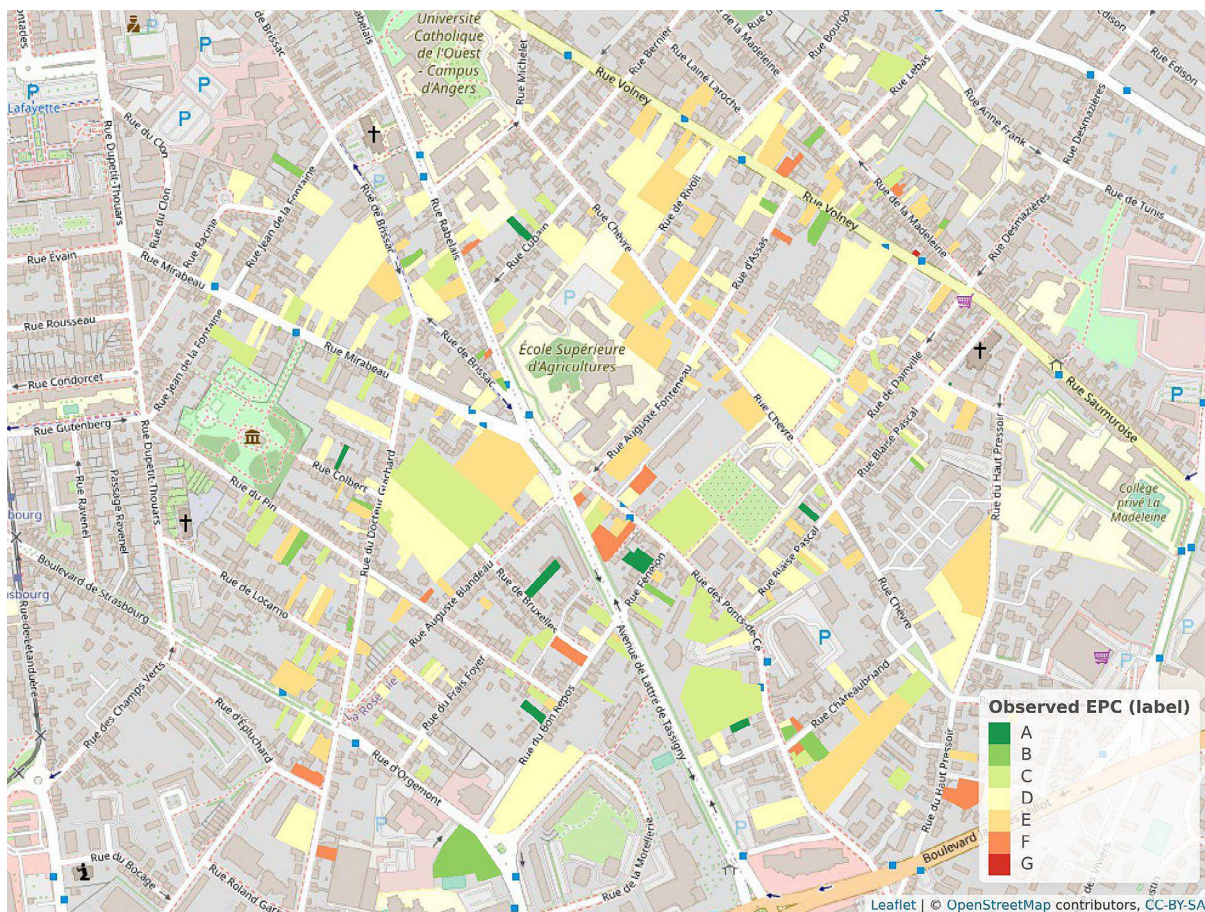


Fig. 4. Map of EPCs in an urban area showing the large amount of missing data.

True values	Predicted values						
	A	B	C	D	E	F	G
A	2	1	1	2	2	0	0
B	1	3	3	9	2	2	0
C	1	3	3	26	15	4	0
D	3	5	5	80	33	5	1
E	4	2	2	36	36	5	1
F	0	3	3	4	5	3	0
G	0	0	0	1	1	0	0

Fig. 5. Confusion matrix of predicted values vs true values using Mixture Kriging. Data comes from a neighbourhood of Angers (France).

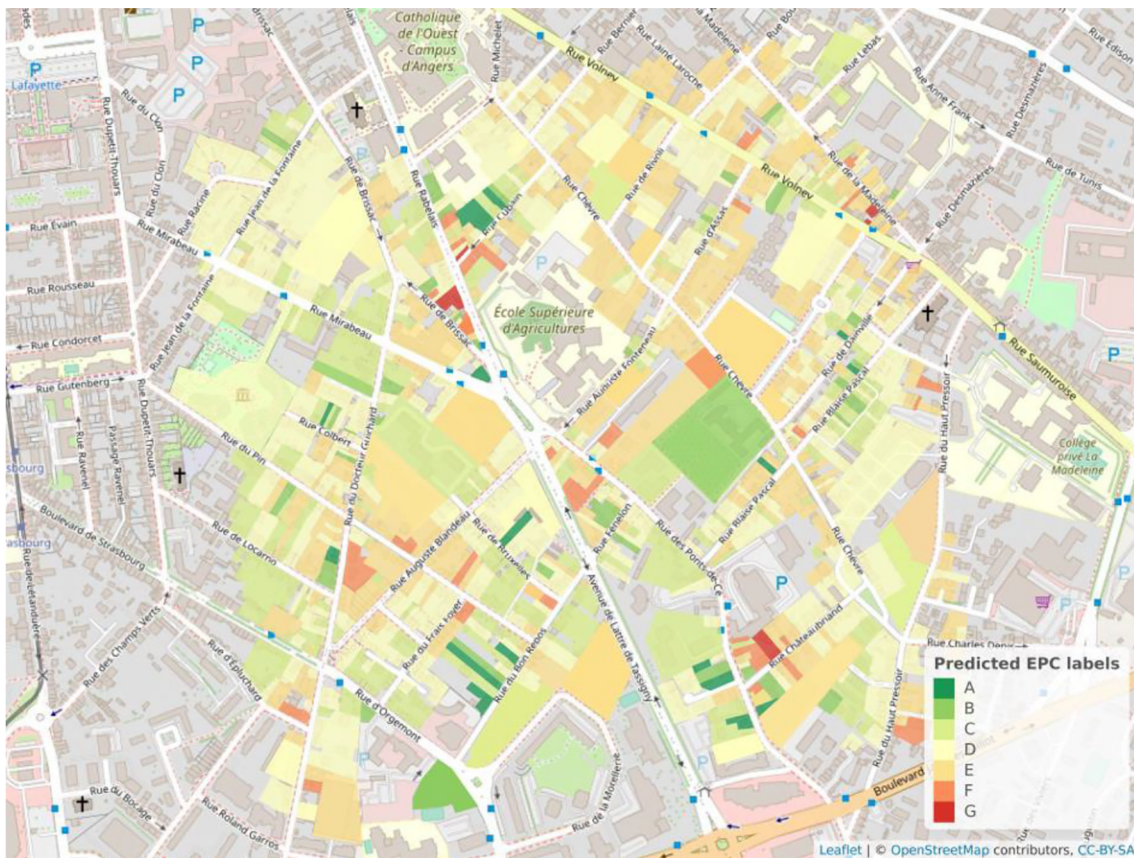


Fig. 6. Map of predicted EPCs in an urban area.

Data availability

Data will be made available on request.

Acknowledgments

The authors acknowledge support from the URBS enterprise, www.urbs.fr. They thank in particular Maximilien Brossard for careful reading and constructive comments.

This research was jointly supported by Mines Saint-Etienne graduate engineering school and research institute (<https://www.mines-stetienne.fr/>

en/), URBS enterprise (<https://www.imope.fr/>) and French National Agency for Research and Technology (<https://www.anrt.asso.fr/fr>).

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Further reading

[1] W.J.K. Aldworth, Spatial Prediction, Spatial Sampling, and Measurement Error, Doctor of Philosophy Iowa State University, 1998.

- [2] U. Ali, M.H. Shamsi, M. Bohacek, C. Hoare, K. Purcell, E. Mangina, J. O'Donnell, A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings, *Appl. Energy* 267 (2020).
- [3] E. Baker, P. Challenor, M. Eames, Future proofing a building design using history matching inspired level-set techniques, *J. R. Stat. Soc. Ser. C: Appl. Stat.* 70 (2) (2021) 335–350.
- [4] I. Ballarini, V. Corrado, F. Madonna, S. Paduos, F. Ravasio, Energy refurbishment of the Italian residential building stock: energy and cost analysis through the application of the building typology, *Energy Policy* 105 (2017) 148–160.
- [5] A. Briz-Redon, A Bayesian shared-effects modeling framework to quantify the modifiable areal unit problem, *Spat. Stat.* 51 (2022).
- [6] Comber, A. and Zeng, W. (2019). Spatial interpolation using areal features: a review of methods and opportunities using new forms of data with coded illustrations *Geogr. Compass*, 13(10). eprint: <https://doi.org/10.1111/gec3.12465>.
- [7] N.A. Cressie, *Statistics for Spatial Data Revised Edition*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, John Wiley & sons inc. edition, New York, 1993.
- [8] L.d.C. Godoy, M.O. Prates, J. Yan, An unified framework for point-level, areal, and mixed spatial data: the Hausdorff-Gaussian Process, 2022.
- [9] P. Goovaerts, Kriging and Semivariogram deconvolution in the presence of irregular geographical units, *Math. Geol.* 40 (1) (2008) 101–128.
- [10] C. Gotway, L. Young, Combining incompatible spatial data, *J. Am. Stat. Assoc.* 97 (2002) 632–648.
- [11] M. Gösgens, A. Zhiyanov, A. Tikhonov, L. Prokhorenkova, Good classification measures and how to find them, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, J.W. Vaughan (Eds.), *Advances in Neural Information Processing Systems*, vol. 34, Curran Associates, Inc., New York 2021, pp. 17136–17147.
- [12] M. Grossouvre, D. Rullière, J. Villot, Spatial interpolation using mixture distributions: A Best Linear Unbiased Predictor, preprint <https://hal.science/hal-0.276127/> 2023.
- [13] E.H. Isaaks, R.M. Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.
- [14] Y. Jin, Y. Ge, J. Wang, G. Heuvelink, L. Wang, Geographically weighted area-to-point regression kriging for spatial downscaling in remote sensing, *Remote Sens.* 10 (2018).
- [15] R. Kerry, P. Goovaerts, I.P. Smit, B.R. Ingram, A comparison of multiple indicator kriging and area-to-point Poisson kriging for mapping patterns of herbivore species abundance in Kruger National Park, South Africa, *Int. J. Geograph. Inform. Sci. (IJGIS)* 27 (1) (2013) 47–67.
- [16] P. Kyriakidis, A geostatistical framework for area-to-point spatial interpolation, *Geogr. Anal.* 36 (2004).
- [17] N.S.-N. Lam, Spatial interpolation methods: a review, *Am. Cartograph.* 10 (2) (1983) 129–150.
- [18] C. Li, Z. Lu, T. Ma, X. Zhu, A simple kriging method incorporating multiscale measurements in geochemical survey, *J. Geochem. Explor.* 101 (2) (2009) 147–154.
- [19] Y.-P. Lin, B.-Y. Cheng, G.-S. Shyu, T.-K. Chang, Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in Chunghua County, Central Taiwan, *Environ. Pollut.* 158 (1) (2010) 235–244.
- [20] P. Moraga, S.M. Cramb, K.L. Mengersen, M. Pagano, A geostatistical model for combined analysis of point-level and area-level data using INLA and SPDE, *Spat. Stat.* 21 (2017) 27–41.
- [21] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*. Industrial Electronics Series, Taylor & Francis, Boca Raton, 2013.
- [22] O.J.R. Pereira, A.J. Melfi, C.R. Montes, Y. Lucas, Downscaling of ASTER thermal images based on geographically weighted regression kriging, *Remote Sens.* 10 (4) (2018) 633.
- [23] L. Poggio, A. Gimona, Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach, *Spat. Stat.* 14 (2015) 4–21.
- [24] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning, MIT Press, Cambridge, Mass, 2006.
- [25] M. Rocas, A. García-González, S. Zlotnik, X. Larráyoz, P. Díez, Noninvasive uncertainty quantification for automotive crash problems with VPS/Pamcrash, *Finite Elem. Anal. Des.* 193 (2021), 103556.
- [26] T. Roksvåg, I. Steinsland, K. Engeland, A two-field geostatistical model combining point and areal observations – a case study of annual runoff predictions in the Voss Area, *J. R. Stat. Soc. Ser. C: Appl. Stat.* 70 (4) (2021) 934–960.
- [27] B.J. Smith, J. Yan, M.K. Cowles, Unified geostatistical modeling for data fusion and spatial heteroskedasticity with R package ramps, *J. Stat. Softw.* 25 (2008) 1–21.
- [28] P. Truong, G. Heuvelink, Bayesian area-to-point kriging using expert knowledge as informative priors, *Int. J. Appl. Earth Obs. Geoinf.* 30 (2013) 2291.
- [29] Q. Wang, W. Shi, P.M. Atkinson, Area-to-point regression kriging for pan-sharpening, *ISPRS J. Photogramm. Remote Sens.* 114 (2016) 151–165.
- [30] C. Williams, C. Rasmussen, *Gaussian processes for regression*, *Proceedings of the 1995 Conference*, The MIT Press, Cambridge, Mass, 1996.
- [31] S.N. Wood, *Generalized Additive Models: An Introduction with R*, Second Edition CRC Press, Boca Raton, 2017.
- [32] E.-H. Yoo, P.C. Kyriakidis, Area-to-point Kriging with inequality-type data, *J. Geogr. Syst.* 8 (4) (2006) 357–390.
- [33] L.A. Zadeh, Toward a generalized theory of uncertainty (GTU) – an outline, *Inf. Sci.* 172 (1) (2005) 1–40.
- [34] X. Zhang, W. Zuo, S. Zhao, L. Jiang, L. Chen, Y. Zhu, Uncertainty in upscaling in situ soil moisture observations to multiscale pixel estimations with kriging at the field level, *ISPRS Int. J. Geo Inf.* 7 (1) (2018).



Marc Grossouvre is a data-scientist in U.R.B.S. company, currently a PhD. student at the National Graduate school of Mines Saint-Etienne. He is working on assessing buildings energy efficiency without physical inspection.



Didier Rullière is Professor of Data Science at National Graduate School of Mines Saint-Etienne in France. He is also an actuary. His research interest includes risk measures, multivariate probabilistic modeling and dependence structures, statistical modeling and spatial models. He holds an actuary diploma, a PhD and a French habilitation to supervise research from the university of Lyon.



Jonathan Villot is a tenured associated professor in environmental and energy management science at Mines Saint-Etienne (GEO department). His research is focused on the evaluation processes of industrial and territorial systems and organisations. Territorial and industrial systems and organisations are addressed in two manners: Developing models and tools for assessment of environmental impacts and pressures on natural resources, to help decision makers by providing relevant information for public decision making; Proposing models and tools for decision support in terms of energy and environmental questions, helping urban policy makers to predict sustainable performances of innovative solutions.