URBANTWIN

From data to decision:
A joint segmentation and rationale regression method for understanding and supporting low carbon technology adoption

ABM4Energy – Copenhagen - 31.03.2025

Matthias Heinrich[a], Hanbit Lee[a], Léonard Léchot[a], Claudia R. Binder[a]

[a] Laboratory of Human Relationships in Urban Systems (HERUS), EPFL, Switzerland

Project context : An ABM to simulate the co-adoption of low carbon technologies

URBANTWIN

- ❖ Improving buildings thermal effiency is urgent to limit climate change and ensure energy resilience.
- ❖ ABM simulation supports the understanding of agent decisions and the development of policy support (Du, 2022)

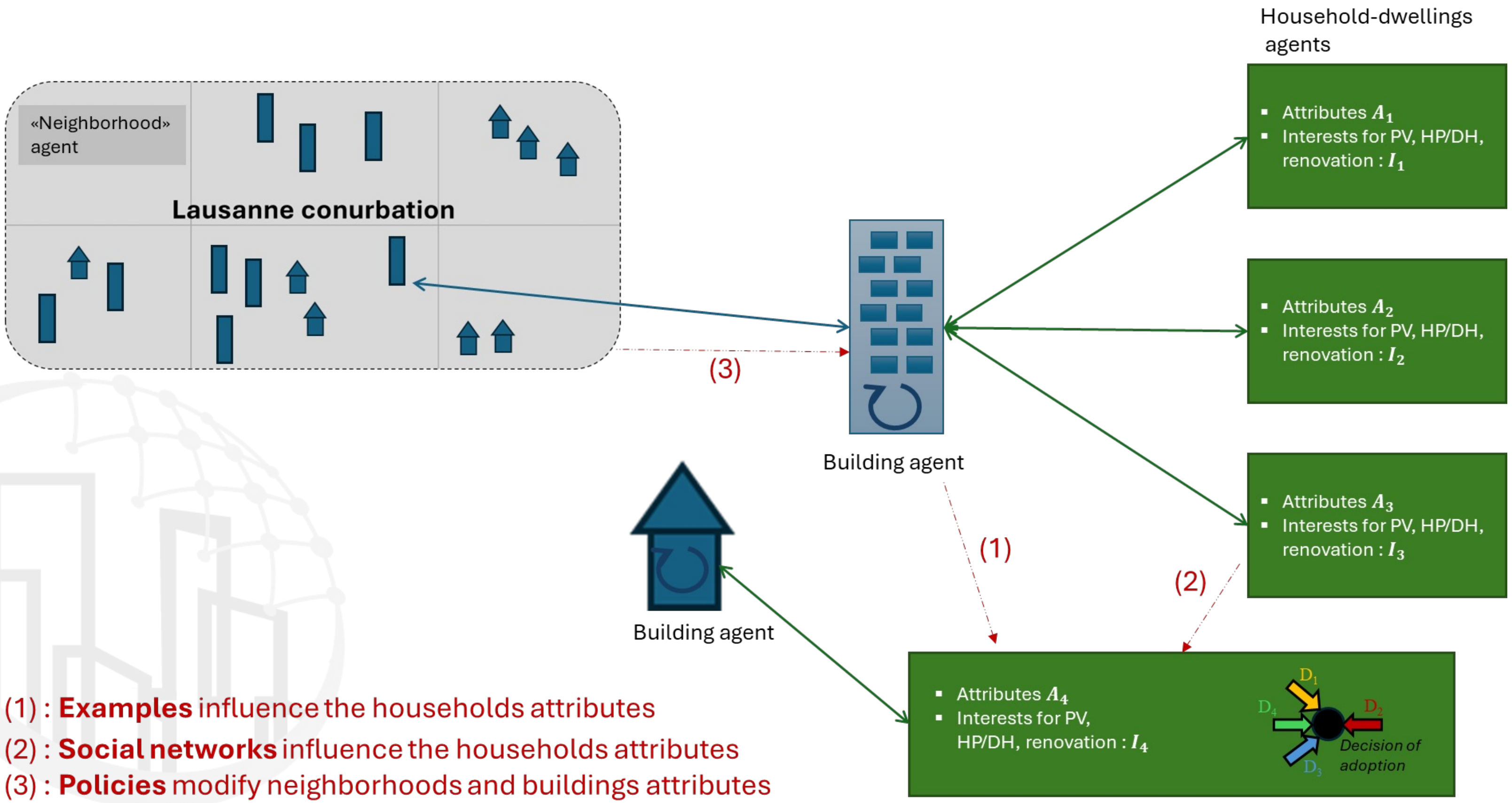
1. Co-adoption How does the adoption of low carbon technologies in buildings influence the adoption levels of other technologies?
2. Optimal policy : Considering these interdependencies, what would be the optimal policy mix?

❖ **UrbanTwin– ETH Board Initiative : Research gaps** EPFL
→ PV, Retrofitting, Heating system replacement in the housing sector
→ Theory-based, spatially explicit and empirical validation
→ Case study : Lausanne region

Theoretical framework
↓
Survey construction
↓
Agent's decision rules identification
↓
ABM development
↓
ABM calibration and validation
↓
Policy analysis

The theoretical framework : Agent types

→ 5 agent types : Environment – Sectors – Building – Dwellings - Households



(1) : **Examples** influence the households attributes

(2) : **Social networks** influence the households attributes

(3) : **Policies** modify neighborhoods and buildings attributes

- Theoretical framework
- Survey construction
- Agent's decision rules identification
- ABM development
- ABM calibration and validation
- Policy analysis

The theoretical framework : Rule definition

Theoretical framework

- **TPB is not enough to model energy retrofit decisions**

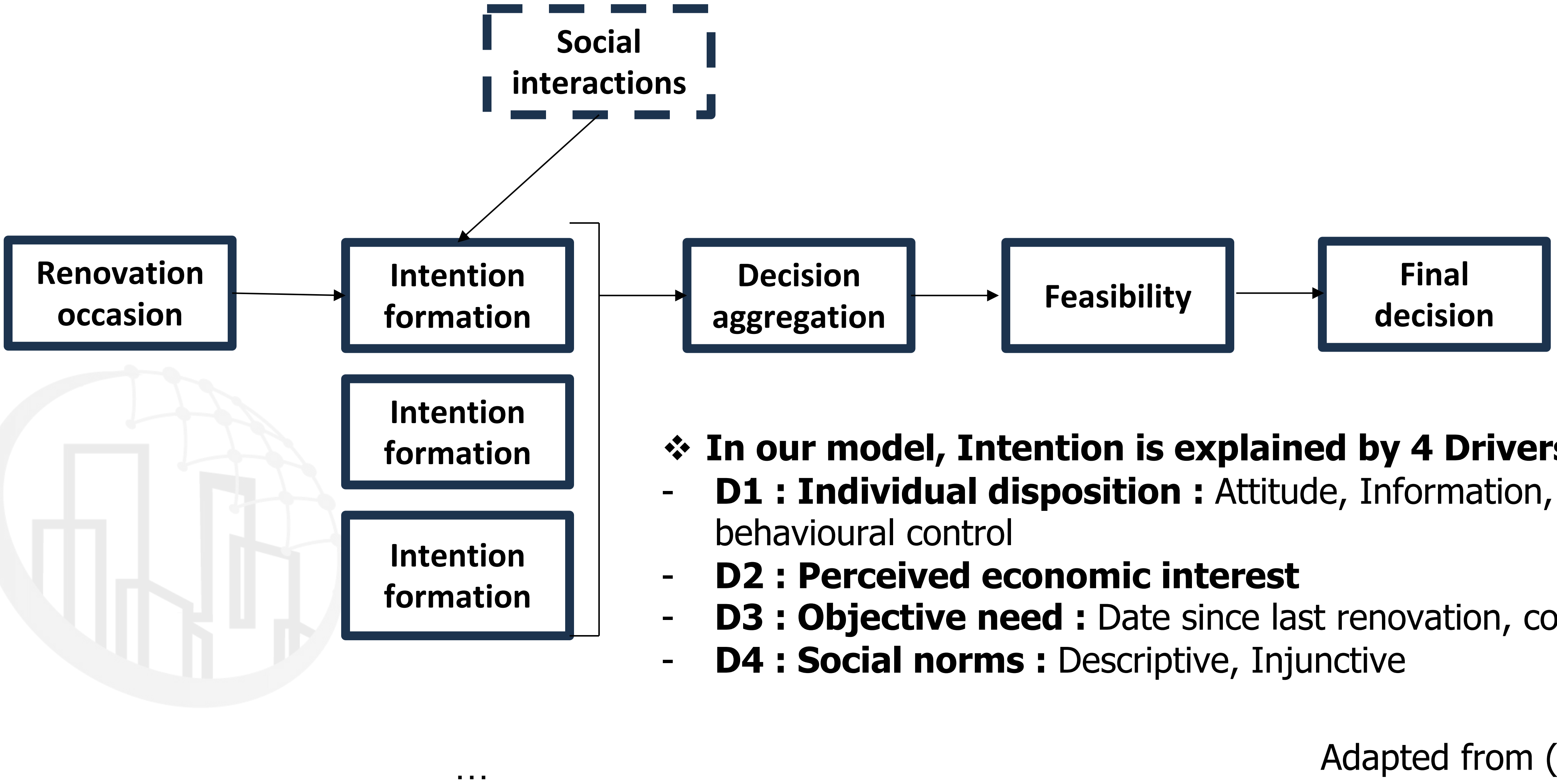
→ Influence of the socio-demographic context (Conradie, 2023)
→ Complexity of the renovation process (Friege, 2016)

- **Our theoretical framework is adapted from the Motivation – Opportunity – Ability (MOA) framework** (Ölander,1995)

- We define agent intention as an ordered variable, based on an adaptation of the model of (Rogers, 1962).

Intention of the household $I_{retrofit,h}$	Description
1	No interest
2	Knowledge
3	Persuasion
4	Decision

ABM conceptual framework



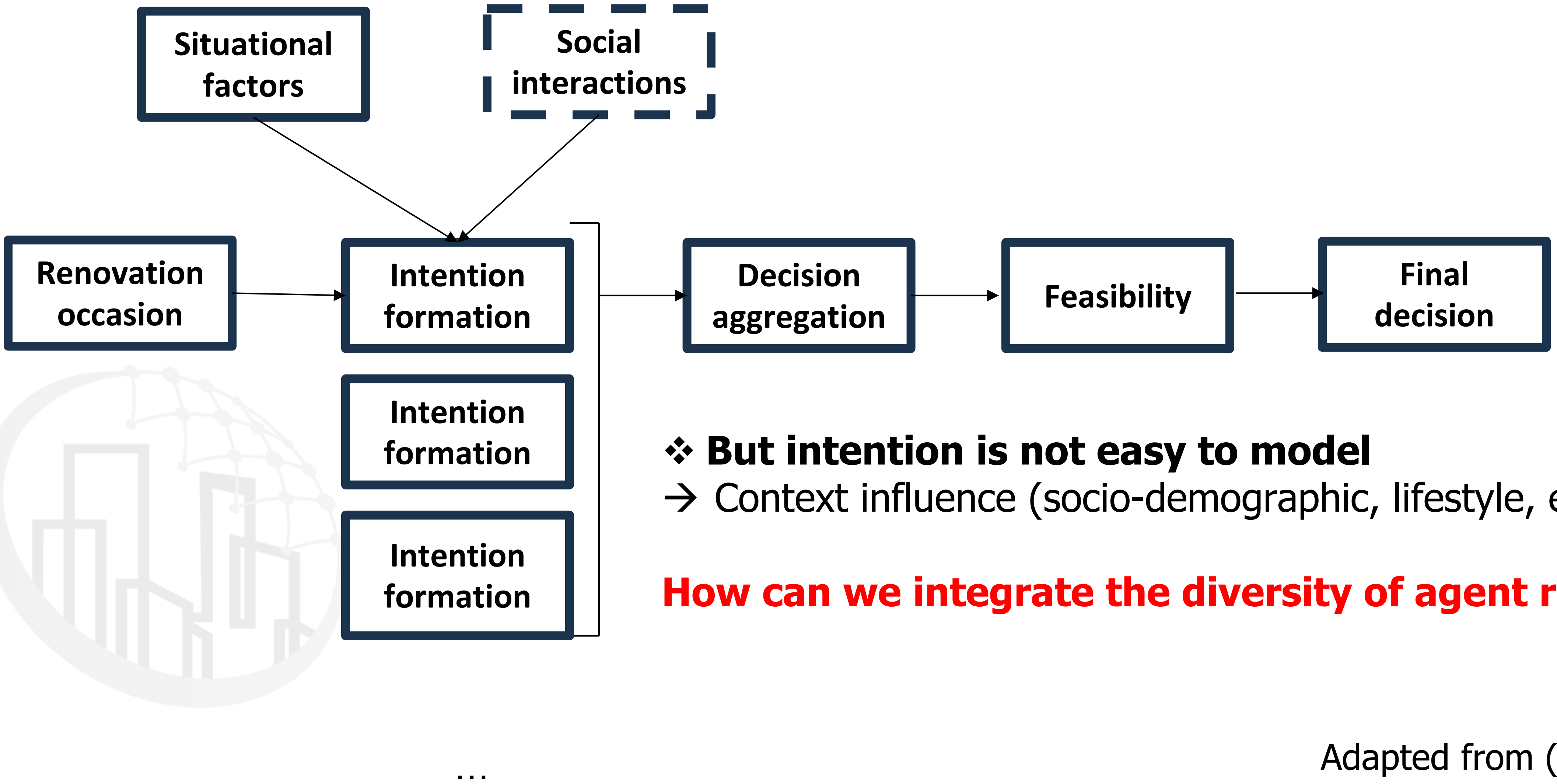
❖ **In our model, Intention is explained by 4 Drivers**

- **D1 : Individual disposition :** Attitude, Information, Perceived behavioural control
- **D2 : Perceived economic interest**
- **D3 : Objective need :** Date since last renovation, comfort
- **D4 : Social norms :** Descriptive, Injunctive

Adapted from (Friege, 2016)

ABM conceptual framework

URBANTWIN



❖ **But intention is not easy to model**

→ Context influence (socio-demographic, lifestyle, etc.)

How can we integrate the diversity of agent rules?

Adapted from (Friege, 2016)

Key insights from the literature on agent heterogeneity

URBANTWIN

An important literature on agent heterogeneity

→ **Clustering agents** : Typology of lifestyles (Frieger, 2016), of households (Sopha, 2013), of buildings
→ **Parameter variability**: Distribution of attributes, Threshold in the utility function (McCoy and Lyons, 2014) (Rai, 2015)
→ **Machine learning** (Zhang, 2021)

❖ **Advantages**

- **Improve model realism**
- **Allow theory exploration**
- **Allow the simulation of emergent patterns of adoption** : Small-scale patterns are driven by agent heterogeneity (Geels, 2017), (Buchmann, 2016)

❖ **Drawbacks and challenges**

- Low **explainability**
- More difficult **calibration and validation** process.
- **Dynamics of the agent rationales** are not always modelled
- **Validation of the clustering/heterogeneity**

Research gap : How to cluster agents so to allow agent heterogeneity, while ensuring explainability, model training, agent rationale dynamics?

Introduction to the model

Data selection

Feature extraction
Driver and synthetic space

Reference model training
Linear model

Advanced model training
Mixture model

The intention model

Notation

I_h : Intention of household *h*
$\boldsymbol{D}_h$: Drivers
$\boldsymbol{\beta}_k$: regression coefficient in the model k

$\boldsymbol{X}_h$: Coordinates of the household *h* in the synthetic space
$\pi_k(\boldsymbol{X}_h)$: weighting coefficient

- ❖ We hypothesize that there are K latent intention models.
- ❖ Each of these intention models considers intention as the weighted sum of independent effects.

$$I_k = \boldsymbol{\beta}_k \cdot \boldsymbol{D}_h$$

- ❖ Let's assume that the households are positioned in a synthetic space $\boldsymbol{X}$
- ❖ We formulate the intention level of each household as the weighted sum of these K models.

$$I_h = \sum_k \pi_k(\boldsymbol{X}_h)\boldsymbol{\beta}_k \cdot \boldsymbol{D}_h$$

Method adapted from (Samé, 2011)

Research questions

- **Research question 1 (RQ1)**
  Is the mixture model a valuable approach for modelling intention compared to a standard linear modelling ?

- **Research question 2 (RQ2)**
  What insights can we gain from this model for policy support?

Introduction to the model

Data selection

Feature extraction
Driver and synthetic space

Reference model training
Linear model

Advanced model training
Mixture model

Data selection

❖ **A survey is being carried out to provide the necessary data**

❖ **The results which are presented here use another data source :**
- Spatial data in Lausanne Region
- 1005 complete responses

Synthetic space	Individual disposition	Perceived eco. interest	Objective	Intention
X_h	D_1	D_2	D_3	I_h
Reference person: Age	Importance of individual changes	Readiness to pay	Duration since the last renovation	Phase of adoption
Household : Number of people, number of children, Date of moving	Need to change to protect the environment	Income level	Thermal performance	
Building : Type, date of construction	Importance of protecting the environment			
	Family expectations regarding environnement protection			

Introduction to the model

Data selection

Feature extraction
Driver and synthetic space

Reference model training
Linear model

Advanced model training
Mixture model

Feature extraction

Feature extraction is used to

- facilitate data representation
- to improve the robustness of the model

X_h : 6 variables	D_1 : 4 variables	D_2 : 2 variables	D_3 : 2 variables	I_h

MDS and MCA are used to reduce the dimension

X_h 2 quantitative var. (40% of the variance)	D_1 1 quantitative var. (14% of the variance)	D_2 1 quantitative var. (16% of the variance)	D_3 1 quantitative var. (17% of the variance)	I_h 1 quantitative var.

Introduction to the model

Data selection

Feature extraction
Driver and synthetic space

Reference model training
Linear model

Advanced model training
Mixture model

Building a reference model

- Linear regression is a widely used method for utility calculation and is used as a reference model

- **Results show**

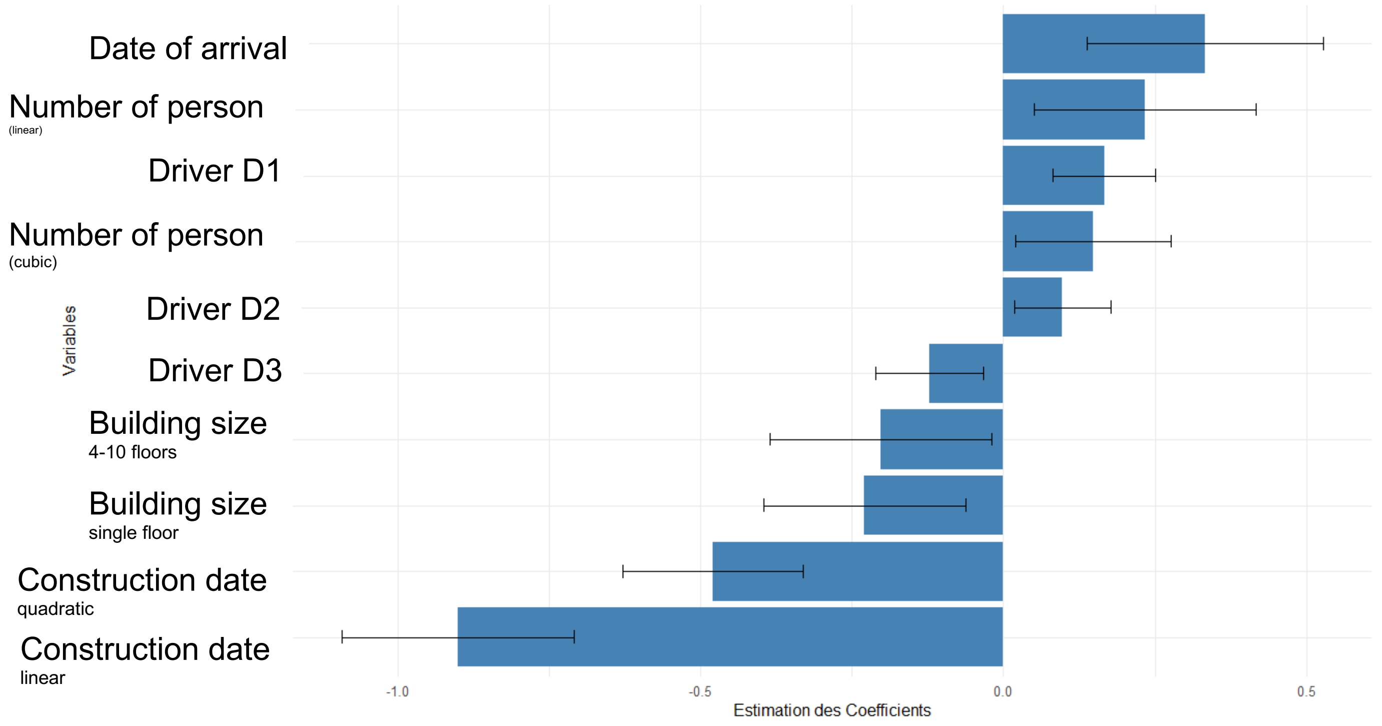
→ Household size positively influence Intention

→ Construction date and building size have a positive and a negative effect on intention

→ D1, D2 and D3 are linked to intention

$$I_h = \boldsymbol{\beta} \cdot \boldsymbol{D}_h$$

Coefficients of the linear regression (significants variables only)



Introduction to the model

Data selection

Feature extraction
Driver and synthetic space

Reference model training
Linear model

Advanced model training
Mixture model

Training the mixture model

$$I_h = \sum_k \pi_k(\boldsymbol{X}_h)\boldsymbol{\beta}_k \cdot \boldsymbol{D}_h$$

Identification of the optimal K

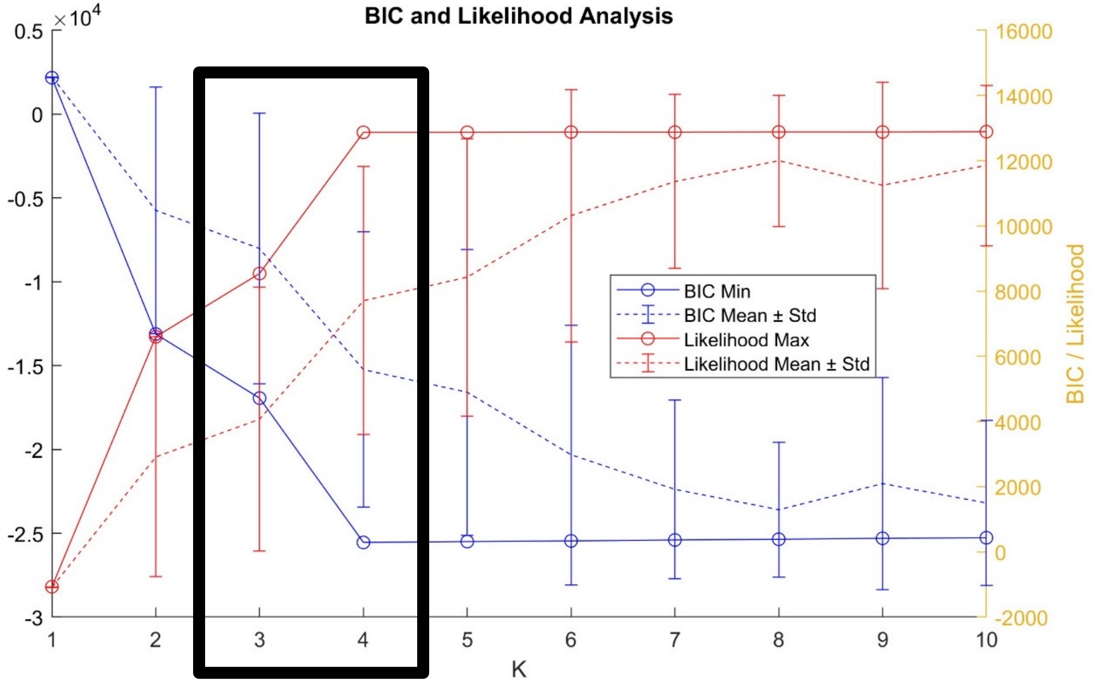
<u>Quantitative criteria :</u>

- Cross validation using the BIC criterion (80% dataset) and the likelihood of the model

<u>Qualitative criteria :</u>

- Cluster homogeneity
- Regression model coherence

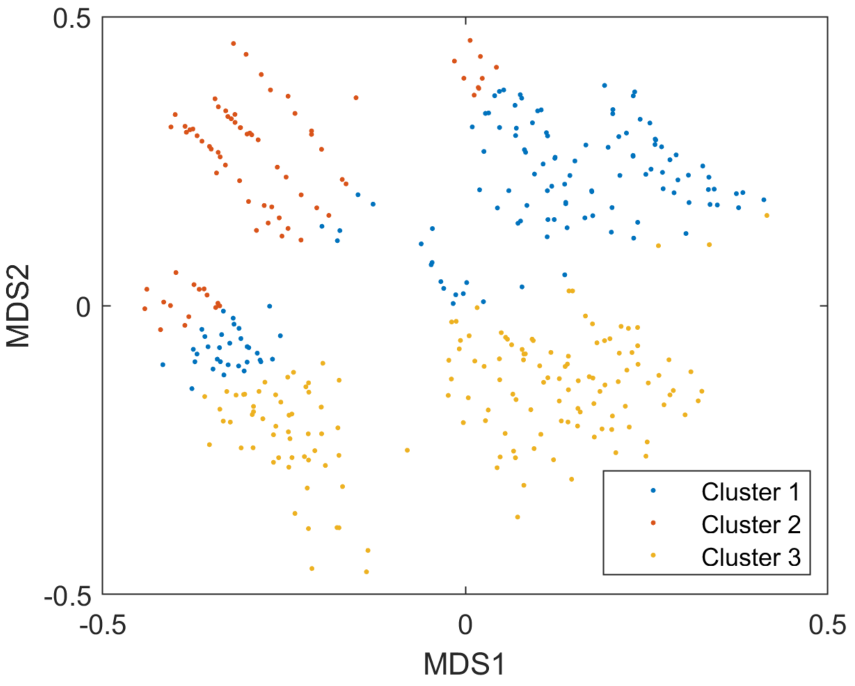
➔ Quantitative indicators suggest that selecting K=3 or K=4 would be a favorable option.



For each K, the model is trained 30 times

Analyze of the mixture model : Composition

❖ **We identify K = 3**



Cluster	Count	Percent
1	307	30.5%

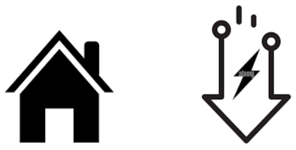
→ *Singles or couples*
→ *Old buildings*

Cluster	Count	Percent
2	171	17%

→ *Recent buildings*
→ *Families*

Cluster	Count	Percent
3	527	52.5%

→ *Old buildings*
→ *Long term residents*



Analyze of the mixture model : Regression coefficients

	Linear model	Cluster 1	Cluster 2	Cluster 3
β_1 Effect of attitudes	**0.16**	0	0	**0.23**
β_2 Readiness to pay	**0.11**	**0.05**	0	**0.05**
β_3 Duration since last renovation	**-0.13**	**-0.9**	**-0.52**	**0.82**

Effects are not the same among the clusters

•**Resident attitudes**: Long-term and older residents exhibit a significant impact of attitudes on their decisions.
•**Willingness to Pay**: Households in the early stages of life, living in high-performing homes, show no willingness to pay for renovation
•**Time since last renovation**:

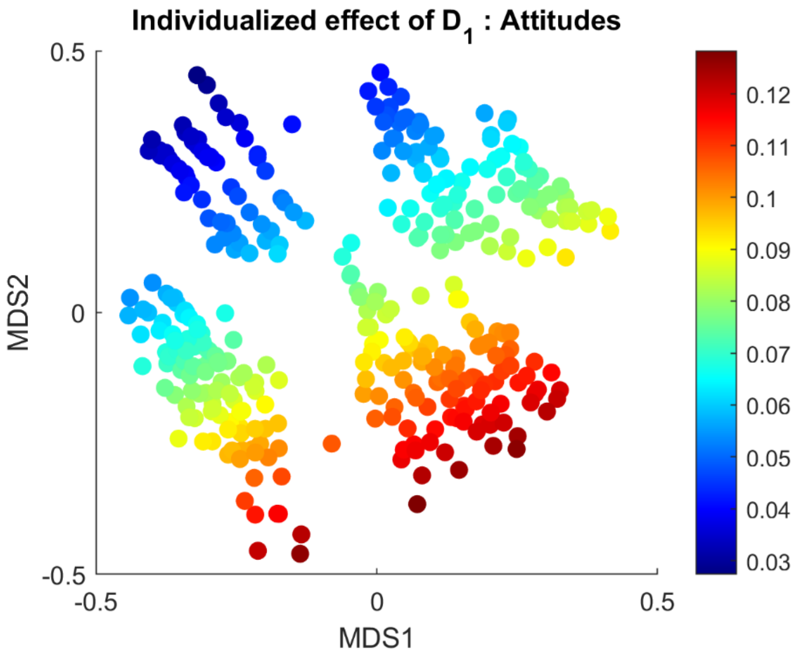
- **Negative** effect for those who have renovated infrequently.
- **Positive** effect for those who have already undertaken renovations.

Conclusion

RQ 1 : Is the mixture model a valuable approach for modelling intention in an ABM ?

- ✓ **An integrated framework** to identify altogether homogeneous clusters and to intention patterns
- ✓ **Multiscale model allowing complexity and interpretability**
  The approach combines type identification and individual level intention modelling

$$\boldsymbol{\beta}_{eq,h,D} = \sum_{k} \pi_k(\boldsymbol{X}_h)\boldsymbol{\beta}_{k,D}$$



Conclusion

RQ 1 : Is the mixture model a valuable approach for modelling intention in an ABM ?

- ✓ **An integrated framework** to identify altogether homogeneous clusters and to intention patterns
- ✓ **Multiscale model allowing complexity and interpretability**
  The approach combines type identification and individual level intention modelling
- ✓ **Dynamic evolution of the rationales** is integrated in the model

Conclusion

RQ 1 : Is the mixture model a valuable approach for modelling intention in an ABM ?

- ✓ **An integrated framework** to identify altogether homogeneous clusters and to intention patterns
- ✓ **Multiscale model allowing complexity and interpretability**
  The approach combines type identification and individual level intention modelling
- ✓ **Dynamic evolution of the rationales** is integrated in the model

RQ 2 : What insights can we gain from this model for policy support?

- ✓ **Empowering long-term residents**: harnessing attitudes to drive sustainable renovations
- ✓ **Incentivizing young households**: unlocking investment in home improvements through financial support
- ✓ **Revitalizing infrequent renovators**: tailored programs to encourage home upgrades and reward commitment

Pending questions

→ Check the validity of the results on another dataset
→ What is the optimal feature extraction strategy ?
→ How could we model intention rules evolution in ABM ?

Contact : matthias.heinrich@epfl.ch

Key references

Hansen, P., Liu, X., & Morrison, G. M. (2019). Agent-based modelling and socio-technical energy transitions: A systematic literature review. *Energy Research & Social Science*, *49*, 41-52.

Tanny, S. M., & Derzko, N. A. (1988). Innovators and imitators in innovation diffusion modelling. *Journal of Forecasting*, *7*(4), 225-234.

Kiesling, E., Günther, M., Stummer, C., & Wakolbinger, L. M. (2012). Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, *20*, 183-230.

Rogers, Everett M. (1962). *Diffusion of innovations* (1st ed.). New York: Free Press of Glencoe.

Friege, J., Holtz, G., & Chappin, É. J. (2016). Exploring homeowners' insulation activity. *Journal of Artificial Societies and Social Simulation*, *19*(1), 4.

Kavak, H., Padilla, J. J., Lynch, C. J., & Diallo, S. Y. (2018, April). Big data, agents, and machine learning: towards a data-driven agent-based modeling approach. In *Proceedings of the annual simulation symposium* (pp. 1-12).

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., ... & Buchmann, C. M. (2016). Simple or complicated agent-based models? A complicated issue. *Environmental Modelling & Software*, *86*, 56-67.

Ren, X., & Chow, J. (2024). *Group Level Agent-Based Mixed Logit for Nonparametric Estimation of K-Modal Taste Heterogeneity with a Ubiquitous Data Set* (No. TRBAM-24-01415).

Geels, F. W., Sovacool, B. K., Schwanen, T., & Sorrell, S. (2017). *The socio-technical dynamics of low-carbon transitions*. Joule, 1(3), 463-479.

Samé, A., Chamroukhi, F., Govaert, G., & Aknin, P. (2011). Model-based clustering and segmentation of time series with changes in regime. *Advances in Data Analysis and Classification*, *5*, 301-321.

Zhang, H., & Vorobeychik, Y. (2019). Empirically grounded agent-based models of innovation diffusion: a critical review. *Artificial Intelligence Review*, *52*, 707-741.