Gall *et al.*
European Transport Research Review (2023) 15:45
https://doi.org/10.1186/s12544-023-00622-0

European Transport Research Review

ORIGINAL PAPER

Open Access

Integrating future trends and uncertainties in urban mobility design via data-driven personas and scenarios

Tjark Gall[1,2*], Sebastian Hörl[2], Flore Vallet[1,2] and Bernard Yannou[1]

Abstract

Urban mobility contributes significantly to greenhouse gas emissions and comes with negative social impacts for various groups, such as limited accessibility to opportunity or basic services. Transitions towards sustainable and people-centred urban mobility systems are paramount. Yet, this is accompanied by various challenges. Complex urban systems are accompanied by high uncertainties (e.g., technological progress, demographics, climate change) which are currently not well integrated. Possible solutions originate from design, policymaking, and innovation, with a widespread disconnection due to non-compatible methods. This paper presents a method to improve the ability to design future urban mobility systems by integrating different approaches for modelling what the future could be and who could be the users. The research question is how diverse future user needs can be integrated in design processes for urban mobility systems. The proposed scenario-based design and personas allows to create data-driven proto-personas—a set of archetypical users with assigned characteristics and behaviours—test their validity, derive distributions across geographical areas, and transform them for different 2030 scenarios. This serves as input to create full personas and synthetic populations as intermediary design objects for the collaboration of designers and simulation experts. The methodology is exemplarily applied in the context of Paris. It contributes to urban mobility solution design that is more aware of future uncertainty and diverse needs of users, therefore, better capable to respond to today's challenges. The approach is replicable with open data and accessible source code: https://github.com/TjarkGall/proto-persona-clustering.

Keywords Inclusive mobility, Persona, Synthetic population, Simulation, Design support

1 Introduction

Urban mobility comes with multiple intertwined challenges, such as rising greenhouse gas (GHG) emissions, unequal access to opportunity, or limited rare materials [1], Climate [2–5]. The United Nations SDG 11.2 emphasises that 'by 2030, [we must] provide access to safe, affordable, accessible and sustainable transport systems for all' [6]. Banister [7] describes the stakes of sustainable urban mobility as more than just the movement of people and goods within cities, but also the interplay between transportation systems, land use patterns, environmental considerations, and social equity. Sustainable urban mobility aims to balance between meeting the diverse mobility needs of urban populations while minimising negative environmental impacts and fostering inclusive and equitable access to transportation.

Achieving this balance can be described as a wicked problem with competing interests, no clear solution, and high levels of uncertainty [8]. Potential solutions are ranging from technologies and urban plans to behavioural nudges and public policies. Planners and

*Correspondence:
Tjark Gall
tjark.gall@urban-framework.com
[1] Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay, 3 Rue Joliot-Curie, 91190 Gif-Sur-Yvette, France
[2] IRT SystemX, 660 Rue Noetzlin, 91190 Gif-Sur-Yvette, France

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

decision-makers use a range of methods to better understand the status quo, the interactions of different components, as well as testing the potential impacts of future solutions, such as new mobility services or public transport expansions. Agent-based simulations are increasingly common to incorporate complex characteristics and interactions between agents, such as congestion, shared mobility use, or rebound effects [9]. Framed in people-centred design, it allows detailed system modelling of impacts on people across time and space [10]. However, two gaps remain: (1) The integration of variations of the future (e.g., population dynamics) in today's design practices, and (2) the disconnection between qualitative and quantitative design, including the transfer of outputs. The resulting research question for this paper is how diverse future user needs can be integrated in design processes for urban mobility systems?

To respond to this question, this article starts by introducing key conceptual components used throughout, namely future scenarios (Sect. 2.1), personas (2.2), and synthetic populations (2.3), followed by an overview of related works. In Sect. 3, the method is described and applied to Greater Paris, followed by discussion and conclusions (Sects. 4–5). The case of Paris is chosen due to the significant role metropolitan areas play in global sustainable urban mobility, the existence of related open-source mobility planning and simulation cases and methods this paper aims to strengthen, and the availability of open data to ensure replicability.

2 Designing future people-centred mobility solutions

This section introduces three existing components that are used in qualitative and quantitative methods to design for future uncertainties (scenarios) and model urban mobility users in predominantly qualitative (personas) and statistically representative (synthetic populations) manner. This section serves the dual purpose of outlining existing works in the field and identifying research gaps, as well as introducing the core components of the methodology described in the following chapter.

2.1 Future scenarios

Scenarios are a tool that permits working strategically with multiple futures. Scenarios are defined as having 'a temporal property rooted in the future and reference external forces in that context […, to] be possible and plausible while taking the proper form of a story or narrative description; and […to] exist in sets that are systematically prepared to coexist as meaningful alternatives to one another' ([11], p. 1). Explorative scenarios are a subset thereof showing different alternative futures to prepare for them today [12]. Future scenarios permit operationalising trends and uncertainties for design and decision-making [13]. Future scenarios have found ample application in urban mobility design (e.g., [14–18]. Most of them follow similar formats which can be simplified via scenario archetypes [15, 19, 20]. So far, few works combine quantitative simulations with explorative future scenarios. While Shaaban et al. [21] apply qualitative scenarios to conduct agent-based simulations, their application and tools are different as the focus is on agri-food systems. Agent-based simulations and scenarios are often mentioned together. However, scenarios are mostly used to define variations of mobility service or, e.g., modal shares, rather than being considered as distinct, possible alternative solutions for the futures. An exception is the paper by Schlenther et al. [22] on simulation-based investigation of transport scenarios for Hamburg. In this case, pre-existing scenarios are compared, focusing on changes of schedule, new mobility services, or, e.g., increased parking pressure simulated via restricted parking plots. Other promising approaches, such as a simulation approach for different potential future street layouts are evolving but with a focus on the spatial dimension [23]. The authors are not aware of an urban mobility simulation where the future population is adapted more than by scaling the population by the average annual growth rate.

2.2 Personas

Personas are a method used in people-centred design that allows considering diversity of people, needs, and lifestyles. They are fictitious characters that represent a homogenous class of users [24]. Future personas allow representing users in a prospective mode and are characterised by (1) a user model, (2) a communication tool; (3) a decision aid and prospective tool [25]. Usually, a persona is given a name, a visualisation, and a narrative that provides input on attitudes and behavioural traits [26]. This allows the persona to come to life and become embedded in the designer's creative process [24, 27]. Methods to create personas can be grouped in three approaches [28]. First, fictional elements can be completed with qualitative data. For example, Goodman-Deane et al. [29] created personas by combining data from questionnaire and interviews. Second, personas can be created from fictional elements [30]. A mixed-method approach is used to create tangible futures and match qualitative mobility user personas with scenarios. Third, recent approaches are using big data leading to data-driven personas (see persona generator, [31]).

Personas create a foundation for discussion between designers and clients and prime them for deeper immersion in varying user profiles that differ from their own socio-economic profile, contributing to more inclusive design [27]. When designing mobility solutions, personas

were used to feed the emotional design of autonomous shuttles [32] or the design of a ride-sharing service [33].

2.3 Synthetic populations

Different to personas, synthetic populations (SPs) aim to statistically reflect a real population [34, 35]. SPs represent reality via individuals, often grouped into households. Sociodemographic attributes are assigned to the individuals. Various approaches have been created to generate these. Some are based on statistical fitting algorithms [36, 37] and generate individuals so that specific attributes such as age or income classes are distributed in line with reference distributions to preserve spatial heterogeneity. Others use disaggregated, sparse data sets such as household travel surveys. These can be described, for example, by using Bayesian networks [38] or Hidden Markov Models [39]. Further, recent approaches join multiple sources of information through machine learning and deep generative modelling [40, 41].

SPs can be useful to model mobility patterns, e.g., to test potential future services or impacts of policy decisions by modifying the schedules of the agents [42], estimating changing GHG and noise emissions [43], or assessing policies [44].

2.4 Existing works

Some works exist that target structured design or simulation of future urban mobility users. For example, Al Maghraoui et al. [45] developed a qualitative approach to model the experience of mobility users. On the other hand, Kamel et al. [46] and Vosooghi et al. [47] created quantitative approaches to measure and integrate preferences, values, or other individual attitudes into data-driven approaches in the mobility context. Conceptual links between personas and synthetic populations have been identified in the past [48] but have not been applied jointly in the context of urban mobility to the authors' knowledge. Building on the existing methodological components and the existing scientific contributions, some promising gaps have been identified. First, most qualitative design approaches are not made to adequately integrate quantitative characteristics. Additionally, many data-driven approaches lack systemic approaches to integrate qualitative trends due to the need to transform the information. Further, a challenge shared across persona and SP-based approaches, only few works have attempted to integrate future changes of users and populations into the approaches. These approaches are either addressing one specific dimension (e.g., population growth) or scaling up the present in some way. The authors are not aware of any work that integrates differences across different user groups in simulation. This paper proposes a mixed-method approach for urban mobility solution design incorporating future uncertainty, constituted of a data-driven persona generation process and a method to integrate qualitative attributes' variation of agents into SPs.

3 Persona clustering and reweighting method

This section introduces the case study, describes the proposed method, followed by a description of data, and the application of the steps.

3.1 Case study of Paris-Saclay

The method is applied to the Communauté d'Agglomeration Paris-Saclay (CPS), an inter-council partnership of 27 municipalities (Fig. 1) south of Paris. It contains an urban hub (Massy), two regional train lines, a planned new metro, and two plateaus, with a linear sub-urban development along one of the regional railways. It borders the Metropolis of Greater Paris (MGP; blue in Fig. 1). The functional mix consists of agricultural uses, low-density mixed-used developments, as well as science and technology clusters. The focus lies on mobility design of CPS for the medium-term future of 2030. However, as a significant relationship exists with Paris, the MGP, and the surrounding areas, such mobility studies should be conducted on the regional scale. The underlying data is the recent 2019 census [49] with information on individual and household level and the national mobility survey [50] from 2018 to 2019.

3.2 Description of 10-step methodology

The method starts with building and validating a set of proto-personas (PPs) based on census data and mobility surveys for the present (Step 1–5). PPs are defined as a pre-stage for personas. They contain a set of quantitative information that are archetypical for mobility behaviours of a part of the population [51]. In the process, PPs are created independently of age and area to create personas that can be used across age groups and living locations. Steps 6–8 compile trends and uncertainties and match them with scenarios, followed by a proportional redistribution of PPs (Step 9). Lastly, the focus is on transforming the generated data into design objects for qualitative and quantitative approaches (Step 10). Figure 2 shows the steps, including the methods and tools used for each step, and the input data. The core steps of the method are fully reproducible with open data and a public repository.[1] In the next part, each step is described in detail. Intermediary results are shown throughout, while final outputs are in the results section.

[1] https://github.com/TjarkGall/proto-persona-clustering.

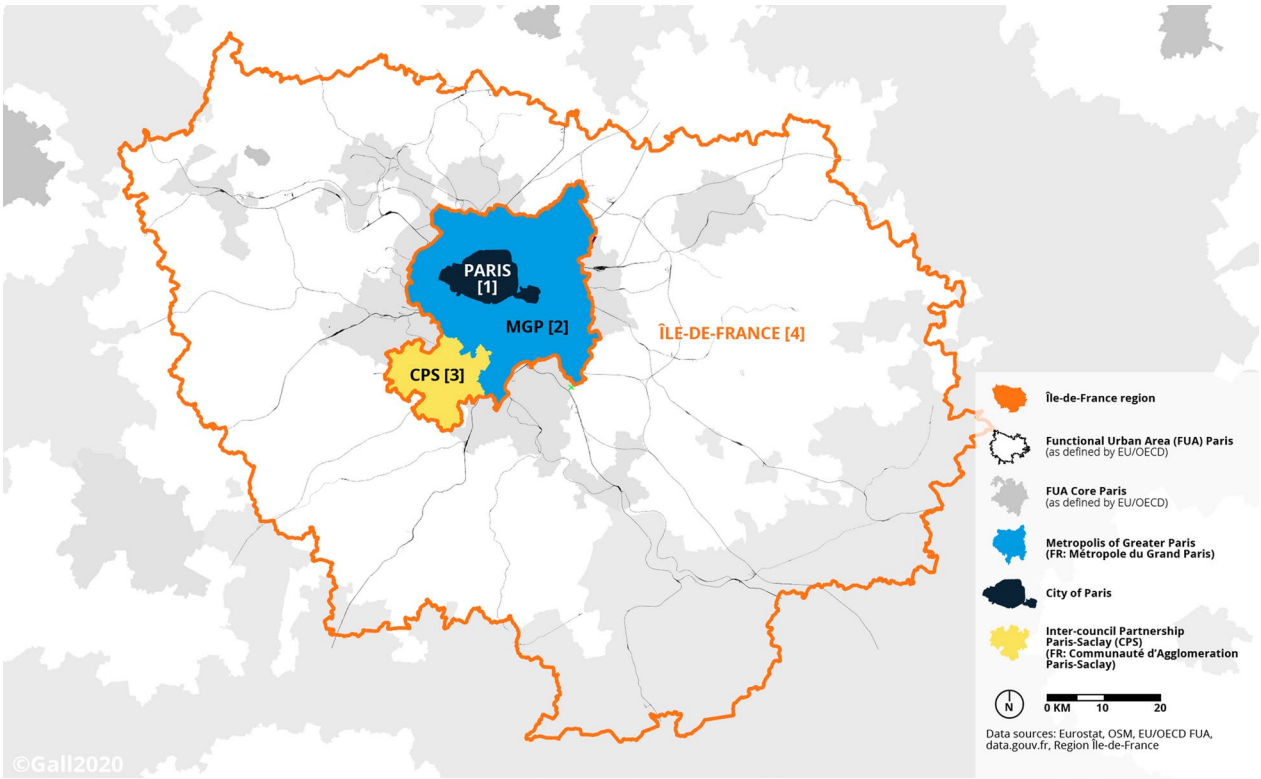

Fig. 1 Four geographical areas used to categorise census data

3.2.1 Step 1: Choosing cluster attributes

Three categories of census data attributes are selected from the available ones: (1) Attributes for clustering (e.g., household size), (2) Attributes for categorisation (i.e., age, area), (3) Attributes to supplement PPs. For the selection of variables, three considerations are important. Some variables can be categorised based on literature or inherent requirements of the process. Second, subjective choices inform the clustering in the context of mobility behaviours and preferences. Third, a verification for partial dependence of several variables supports the choice of most suitable variables. 25 variables are the basis for the following data preparation and clustering.[2]

3.2.2 Step 2: Data preparation

The selected attributes are organised, their data types transformed, and individual entries scaled stochastically by their statistical weight to a total of nearly 14 M individuals. Attributes that are not considered as relevant are deleted. To account for the geographical focus and mobility flows between the focus area CPS and its surroundings, geographical codes are assigned for each record. To avoid overrepresentation and complexification, a statistical analysis is performed to test for correlations between numerical variables aiming to reduce the number of attributes if some of them are significantly correlated. Statistically significant ($p < 0.05$) and high degree correlations exist between the number of children under 17 (NE17FR) on the one side and children under 5 (NE5FR) as well as household size (NPRR) on the other. Consequently, NE17FR is excluded from the clustering. For the remaining combinations, no variable can be identified which explains another variable strongly enough that exclusion would be justified.

3.2.3 Step 3: Proto-persona (PP) clustering

Next, census entries are clustered based on selected attributes. The mix of categorical (e.g., socio-professional category) and numerical info (e.g., number of cars) excludes common clustering approaches. The k-prototype method is used which was developed to cluster including categorical data [52]. The method searches most suitable central values for a set number of clusters. These central values are adapted over n iterations to improve the model fit. The quality of the outcome can be defined by the distance between the values

[2] Tables 8 and 9 in the Annex show selected attributes and codes from the 2019 Census. Descriptive statistics are provided in Tables 10, 11, 12.

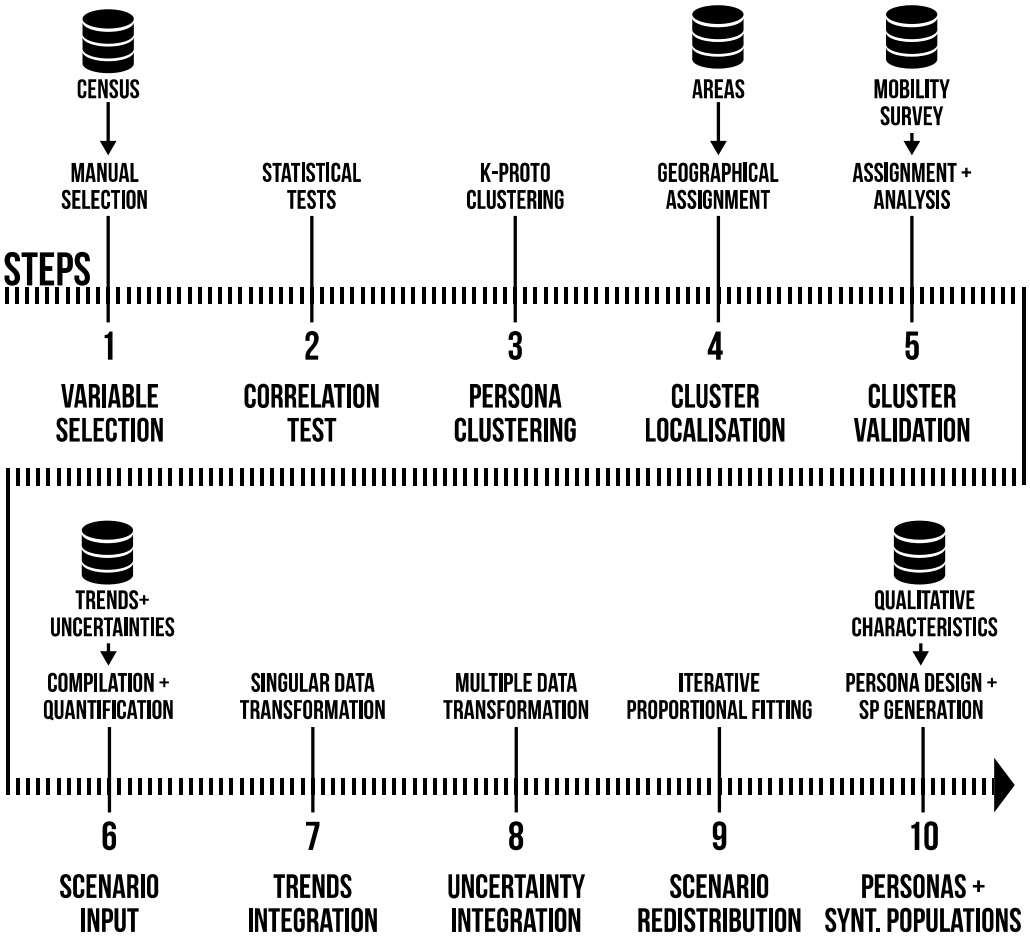

Fig. 2 Combined development process of data-driven personas and synthetic populations for future scenarios

assigned to each cluster and the central values. For example, if the number of cars per household of one cluster is one but some data entries with two cars are still assigned to the cluster due to a high fit in other categories, this difference is defined as distance to be minimised.

To cluster, the number of clusters is defined. This can either be a subjective choice or the result of a preceding test. For the latter, a scree-test can be performed [52]. The objective function is made up from the 'within-ness','the 'sum over all clusters within distances to the prototypes for each cluster [52], p. 202). By running the clustering process for a variety of cluster sizes (k = 2, 4, […] 50) with 100 iterations, the scores can be compared. This process is conducted with a 5% sample of the overall scaled population data (n = 683 k). While more clusters

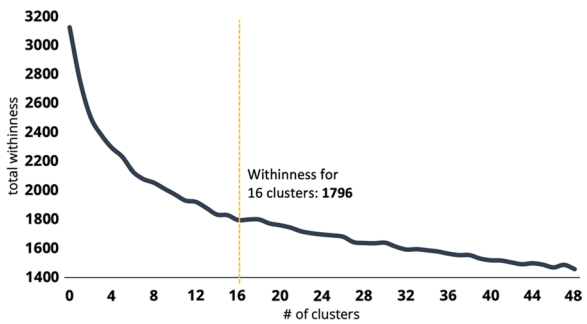

Fig. 3 Scree test graph showing total within-ness (y-axis) for different number of clusters (x-axis)

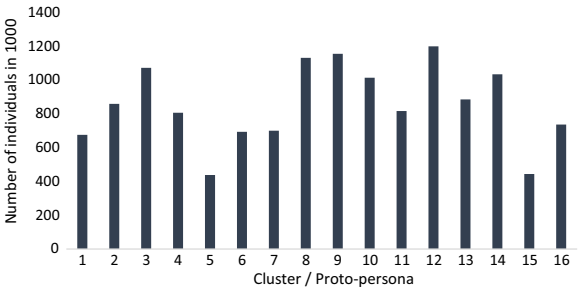

Fig. 4 Distribution of entries by cluster

Table 1 Selected central values for 16 clusters / proto-personas

PP	Rooms	Child < 5	Houshold	Car	Couple		Degree	Student	Parking	Activity locaction	Birthplace
					Socio-prof						
1	4	1	4	1	Yes	Without job		Yes	Yes	In municip.	In dep.
2	4	1	4	1		Higher ed.	Master			Outs. dep.	Abroad
3	4		2	1		Retired	Vocational		Yes		Outs. region
4	2		1	1	Yes	Higher ed.	Master			Outs. dep.	Outs. region
5	4	1	4	1		Employee	Vocational		Yes	Outs. municip.	Abroad
6	3		3	1	Yes	Without job		Yes		In municip.	In dep.
7	5		4			Higher ed.	Master		Yes	Outs. dep.	Outs. dep.
8	4	1	4	2	Yes	Without job					In dep.
9	3		3	1	Yes	Employee	Vocational			Outs. dep.	Outs. dep.
10	2		1		Yes	Higher ed.	Master			In municip.	Abroad
11	4	1	4	1		Without job	Vocational		Yes		Abroad
12	4		2	1		Higher ed.	Master		Yes	Outs. dep	Outs. region
13	5		3	2	Yes	Intermediary	Technician		Yes	Outs. municip.	In dep.
14	5		4	2	Yes	Without job	College	Yes	Yes		Outs. dep.
15	2		2		Yes	Without job	Highschool				Abroad
16	3		1	1	Yes	Retired	Vocational				Outs. region

PP	Way of life	Gender	Status		Housing	Size sqm	Activity	Empl. time	Transport	Type
1	Child of couple	F	Single		Owner	60–80	< 14 years			Apartment
2	Couple with children	M	Married		Owner	80–100	Employed	Full time	Public transport	Apartment
3	Adults without children	M	Married		Owner	80–100	Retired			House
4	Adults	M	Single		Tenant	40–60	Employed	Full time	Public transport	Apartment
5	Couple with children	F	Married		Owner	60–80	Employed	Full time	Car	Apartment
6	Child of single parent	M	Single	Tenant social housing		60–80	< 14 years			Apartment
7	Couple with children	F	Married		Owner	120 +	Employed	Full time	Car	House
8	Child of a couple	M	Single	Tenant social housing		80–100	< 14 years			Apartment
9	Single parent	F	Single	Tenant social housing		60–80	Employed	Full time	Public transport	Apartment
10	Single under 40	F	Single		Tenant	< 30	Employed	Full time	Public transport	Apartment
11	Couple with children	F	Married	Tenant social housing		60–80	Unemployed			Apartment
12	Adult without children	M	Married		Owner	60–80	Employed	Full time	Car	Apartment
13	Child of a couple	M	Single		Owner	80–100	Employed	Full time	Car	House
14	Child of a couple	M	Single		Owner	120 +	Students			House
15	Single under 40	M	Single		Tenant	< 30	Unemployed			Apartment
16	Single over 40	F	Widow		Owner	40–60	Retired			Apartment

are usually producing more accurate results, the goal is finding a balance between a number as low and an accuracy as high as possible. For this, the elbow method can be applied in which the number of clusters is set at the point where the curve becomes less steep (Fig. 3). Here, 16 clusters are chosen.

Next, all remaining variables are used to cluster the scaled population (n = 13,658,311). This is done with 100 iterations and including non-applicable variable entries (e.g., no indication where job is located as person goes to school). This results in the distribution of entries by cluster shown in Fig. 4. The individuals are grouped in clusters of relatively homogenous characters, between 400 k and 1.2 M people per cluster.

The clusters represent PPs. A selection of the central values of the cluster outcomes are shown in Table 1. The Table 1 shows the simplified meaning behind the majority of attributes that have been used for clustering, including information concerning the household (e.g., number of children and household members), individual socio-demographics (e.g., degree, activity location, origin), as well as supplementary information (e.g., parking spot

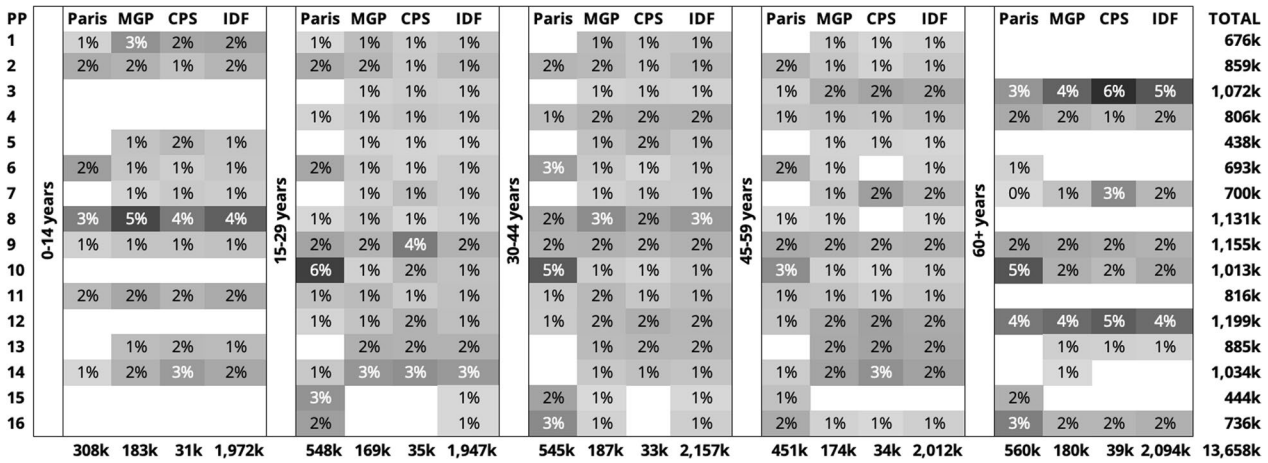

Fig. 5 Cluster occurrences per area and age group in percentage (100% for each area, showing values from 0.5% upwards)

availability or housing type). To test the accuracy of the clusters, the most frequent clusters are chosen to compare its cluster central values and values of all entries assigned to the cluster. The numerical variables show a high accuracy across the variables. For the categorical data, some have very high accuracies and indicate that nearly 100% of data entries share the central value of the PP. In others, values under 50% appear. However, these are primarily in variables with many categories and all values are sufficiently high to evaluate the cluster results as solid and, with the limitation inherent to clustering over 13 million entries into 16 PPs, representative.

These 16 clusters represent groups of people with shared mix of attributes linked to the way of life and mobility behaviours. As an example, PP16 (n = 736 k) is a widowed woman living alone in a 3-room, 40–60 sqm apartment in Paris. She owns her apartment in a multi-unit building but does not own a car. She holds a vocational degree and is retired. Originally from another metropolitan region in France, she moved to Paris. This sample PP shows that with the performed process, comprehensible descriptive profiles can be generated.[3]

3.2.4 Step 4: Proto-personas per age group and area

The clusters represent PPs which do not appear equally often across age groups and area. The objective was developing personas independently of age and area to focus on socio-economic characteristics and introduce these values via scenario variation. However, these variables can be re-associated and analysed. The distribution is organised by cluster, age group (0–14 years, 15–29 years, 30–44 years, 45–59 years, 60 + years) and geographical areas (Fig. 5). This provides an idea in how far all personas can describe sufficient varying classes and represent groups such as children. For example, for the age group of 0 to 14 years, ten clusters have occurrences, with PP8 being the strongest with occurrences between 3 and 5%. While less personas appear in the under 15 and over 60 years group, sufficient PPs are included in both to represent a diversity across the age group.

3.2.5 Step 5: PP validity test for mobility

A challenging element in persona development is validating their accuracy. Even when working with a PP without qualitative attributes, it must be validated that choices of attributes and clustering approach can lead to personas that represent variations in urban mobility. Two approaches can be used for this. First, a subjective visual approach allows to validate the representativeness of the underlying data. If considered relevant, specific personas can be added, for example, underrepresented personas such as people with reduced mobility. Second, mobility surveys can be used to test if statistically significant differences in mobility behaviours across the PPs can be identified. In this paper, this is tested with the mobility of people survey [50] which details what mobility-related resources are available (e.g., driving license, car) and what trips are done where and when on a given day. All entries

Table 2 Information from mobility survey [50]

Variable	Description	Analysis
nb_dep	Number of trips per day	Sum
DUREE	Duration of the trip	Mean, max, sum
MTEMPSMAP	Walking time during the trip	Mean, sum

[3] The reader can perform the exercise for the remaining clusters on the basis of Table 1 (central PP values), Table 8 (variables), and Table 9 (codes for categorical variables).

Table 3 Compilation of trends

Trends	Description	Values		
		Area	Growth rate (%)	
Annual population growth	The population is assumed to continue to grow. The central (incl. migration) scenario [53] is considered for Paris and its environments (equalling petite-couronne and MGP). The IDF data [56] is adapted by overall population to IDF without Paris and MGP	Paris	0.4	
		MGP	0.6	
		IDF/CPS	0.13	
		Age group	**Female**	**Male**
Ageing	The demographic ageing process leads to an increase in the age group of 60+. Different values are calculated based on statistical assumptions from INSEE [53]	0–14	0.00	0.02
		15–29	−0.07	−0.08
		30–44	0.00	−0.14
		45–59	−0.03	−0.02
		60+	1.65	2.50

Table 4 Compilation of uncertainties

Uncertainties	Description	Range	Sources
U1: Car ownership	Car ownership, in particular in Paris, is assumed to decrease, while in more remote places an increase is possible. Changing technology, prices, and policymaking allows various assumptions	Stable to less inside Paris/MGP, less to more outside MGP	[57]
U2: Work/study location	With increasing transportation speed and trends like working-from-home, more people might work further away. Congestion or concepts like the 15-min city or urban villages could counteract	Across all regions significant possible changes	[58]
U3: Household size	Household sizes are assumed to decrease. However, its level is unknown	Decreasing, strongest inside Paris	[59]

are considered where both origin and destination of trips were inside the Île-de-France region. The minimum, maximum, sum, and average values for variables (aggregated from 0 to 8 trips) are calculated. Table 2 shows the input variables, their description, and the performed calculations. For this paper, three variables are chosen.

To match mobility data with the PP, shared variables are used.[4] The results show that differences between categories are in line with expectations. To exemplify this analysis, a random data entry that is matched to PP16 is chosen. A selection of the assigned information is: 1 person, female, single, retired, 1 car, 2 trips, first starting at 11:45:00, average duration 15 min, with a two-hour activity in between, and a 1 km distance. This fits the prior description of PP16. After repeating this process multiple times, it can be concluded that the PPs describe successfully varying mobility behaviours. However, the matching process on few shared variables, as well as the exclusion of under 16-year-olds due to lack of data bear potential for future improvements.

3.2.6 Step 6: Uncertainty/trend compilation

The last steps described preparatory steps with today's data. To integrate possible futures via scenarios, uncertainties and trends can be used to redistribute PPs. Trends are developments that are assumed to materialise. For example, a continuing ageing of the society, a stable male/female distribution, or a population increase for an area. Uncertainties are contextually relevant attributes that are assumed to change but for which it is impossible to be certain of the direction of change. This might be the level of private car ownership or the willingness to take public transport. Each of them can vary. For example, the car ownership might range between 0.8 and 1 per person based on socio-economic and political developments (currently 1 on average). Some key trends (Table 3) and critical uncertainties (Table 4) are chosen for this study.

[4] I.e., household size (± 1), gender, activity type (TACT), and number of cars (± 1).

Table 5 Annual growth rates in percent for each category, based on [53, 56] (Attention: Values do not refer to Ile-de-France but IDF without Paris, MGP and CPS)

Age groups	0–14 y		15–29 y		30–44 y		45–59 y		60 + y	
Gender	M	F	M	F	M	F	M	F	M	F
Paris	+ 0.02	± 0	−0.08	−0.07	−0.14	0	−0.02	−0.03	+ 2.5	+ 1.7
MGP	+ 0.03	+ 0.01	−0.07	−0.06	−0.13	+ 0.01	−0.01	−0.02	+ 2.5	+ 1.7
CPS	+ 0.02	± 0	−0.08	−0.07	−0.14	+ 0.01	−0.02	−0.03	+ 2.5	+ 1.7
IDF	+ 0.02	± 0	−0.08	−0.07	−0.14	± 0	−0.02	−0.03	+ 2.5	+ 1.7

Table 6 Four archetypical scenarios [15]

Scenario	Description
S1: Grumpy old transport	No significant changes to the current status
S2: At an easy pace	Slow, consistent change across fields
S3: Mine is yours	Strong shift towards a shared economy and mobility system (incl. MaaS)
S4: Tech eager	Technology as driving force for majority of future mobility

Table 7 Trends and uncertainties quantified and adapted by authors for each scenario

Scenarios	U1: Car ownership	U2: Proximity of work/study locations	U3: Household sizes
S1	Increase (+ 5%)	Stable (± 0%)	Stable (± 0%)
S2	Decrease (−5%)	More local (± 1.5%/3%)	Decrease (−5%)
S3	Decrease (−5%)	More local (± 2%/4%)	Increase (+ 10%)
S4	Stable (± 0%)	Less local (−/ + 1.5%/3%)	Strong decrease (-10%)

3.2.7 Step 7: Future trend integration

These key trends are integrated depending on PP, age-group, and timescale. An anticipated annual population growth rate can be applied to each of the PPs via their respective occurrence rate (weight) that results from the clustering. On the other hand, an ageing society would lead to a proportional shifting of PPs by age-group. Based on the population and ageing forecasts from the statistical service of France [53], growth rates (Table 5) are calculated which are equally applied to all PP within one area. The growth rate is calculated as combined value of both trends.

3.2.8 Step 8: Uncertainties' values per scenario

The key difference to step 7 is that uncertainties are not applied equally but dependent on an underlying scenario. A number of different possible scenarios (usually 4–7) is defined for a particular year in the future, in this case 2030 (Table 6). Each scenario can be complemented with a name, narrative, and set of associated, application-dependent information. For the use case of this paper, three uncertainties and a numeric adaptation for each is defined. For their adaptation, scenario archetypes can be used [19, 54], in particular those of the mobility sector [15]. Building on these scenarios, values for each uncertainty are set to create coherent scenarios and a wider distribution across the variables (Table 7).

3.2.9 Step 9: Redistribute PP by scenario

Three options exist to adapt the PP for each future scenario. The distribution of the PP can be adapted to best match the values provided for the uncertainties. Second and if necessary, certain attributes of the present PP can be adapted to better allow the matching of future uncertainties. Third, some or all of PPs could be adapted differently for the scenarios. The latter would lead to a much higher number of personas, defeating its purpose. The focus of this paper is on the former option as it is deemed the optimal case for the methodology. Two steps need to be conducted. First, the growth and demographic change scales the overall distribution. The growth rates (Table 5) are applied to scale the PPs without changing the proportion. This constitutes the basis for all scenarios. Next, changes of certain impacts

shall be reflected as accurate as possible by changing the distribution of personas. This is done by optimising the initial population weights so that uncertainty-based target values are reached in the future scenarios. The target scenarios are characterised by target mean values of numerical attributes across the overall population and anticipated target shares of persons with specific attribute values for nominal attributes. The following formula outlines the inputs and basis for the fitting process:

Persona variables and constants

- $K \in \mathbb{N}$: Number of personas
- $\omega'_k \in (0,1)$: Initial population share per persona k(baseline)
- $\omega_k \in (0,1)$: Population share per persona k (to be chosen)

Attribute constants

- $\mathcal{A}$: Set of attributes with a mean target value
- $a_k \in \mathbb{R}$: Value of attribute A in persona k
- $\widehat{A} \in \mathbb{R}$: Target mean value for attribute A
- $\mathcal{B}$: Set of attributes with a target share per value
- $\mathcal{V}_B$: Set of values that are permissible for attribute B
- $b_{k,v} \in \{0,1\}$: Defines whether attribute B in persona k has value v
- $\widehat{B}_v \in \{0,1\}$: Target share for value v of attribute B

The target share values need to fulfil:

$$\sum_{v \in \mathcal{V}_B} B_v \leq 1. \forall B \in B$$

If target shares for all possible values are given, equality must hold.

Optimisation constants

- ξAB : Objective weight for attribute AB
- α : Regularisation weight

Problem formulation

$$\begin{aligned} \min_{\omega_k} & \sum_{A \in \mathcal{A}} \xi_A \left(\hat{A} - \sum_k \omega_k a_k \right)^2 \\ & + \sum_{B \in \mathcal{B}} \sum_{v \in V_B} \xi_B \left(\hat{B}_v - \sum_k \omega_k b_{k,v} \right)^2 \\ & + \alpha \sum_k \left(\omega_k - \omega'_k \right)^2 \\ \text{s.t.} \quad & \sum \omega_k > 0 \end{aligned} \tag{1}$$

The first term in the objective calculates the mean of an attribute a given the updated weights and compares it to the required target value. Similarly, for nominal attribute values where a target value has been defined, the second term compares the obtained share of all PPs that have this specific value with the requested share. Each objective (representing an attribute or an attribute value) can be prioritised through a weighting factor. For attributes with higher attribute weights, matching their target value will be enforced with higher importance than for other attributes. Finally, a regularisation term is introduced which allows to give importance to not deviating too strongly from the initial weights of each persona. The regularisation weight defines whether the weights will be adapted in an aggressive way (increasing few personas strongly) or more uniform (increasing many personas carefully), but also has an impact on how well the posed target values can be attained. In the present use case, the number of cars has a weight of 0.4, household size 0.3, work/study location 0.3 and a weight of 0.1 was given to the regularisation term to keep the variation between initial and final weights as small as possible. The outcomes are provided in the results section.

3.2.10 Step 10: Create personas and synthetic populations (SPs)

Finally, present and future sets of PPs and their distributions are used to create full personas as intermediary design objects. Simultaneously, PPs distribution by area, in combination with travel surveys, and registries of residential, work, commercial, and leisure activities can be used to generate one present, and a set of future SPs. Building on the additional information assigned from the mobility survey to the personas (e.g., most common activity chains), qualitative persona creation methods allow building a set of personas (incl., for example, name, photo, location [30]).[5] This is described by way of example in the results section. The difference to other processes is that the starting point (age, characteristics, location, activity chain) is not defined by the designer but instead comes from the underlying evidence-based approach and the supplementary information. Further, building on Vallet et al. [48], we can use the generated distributions and profiles for the generation of multiple synthetic populations. The only component that needs to be changed in the synthesis process are the initial weights in the census information. As, by using the clustering algorithm, a persona can be assigned to each record in the census data, their weights can be adapted according

[5] A full set of personas based on this approach are available online: https://urban-mobility-futures.notion.site/Personas-aa3b30f47c354220bc025dd7edb207cd?pvs=4.

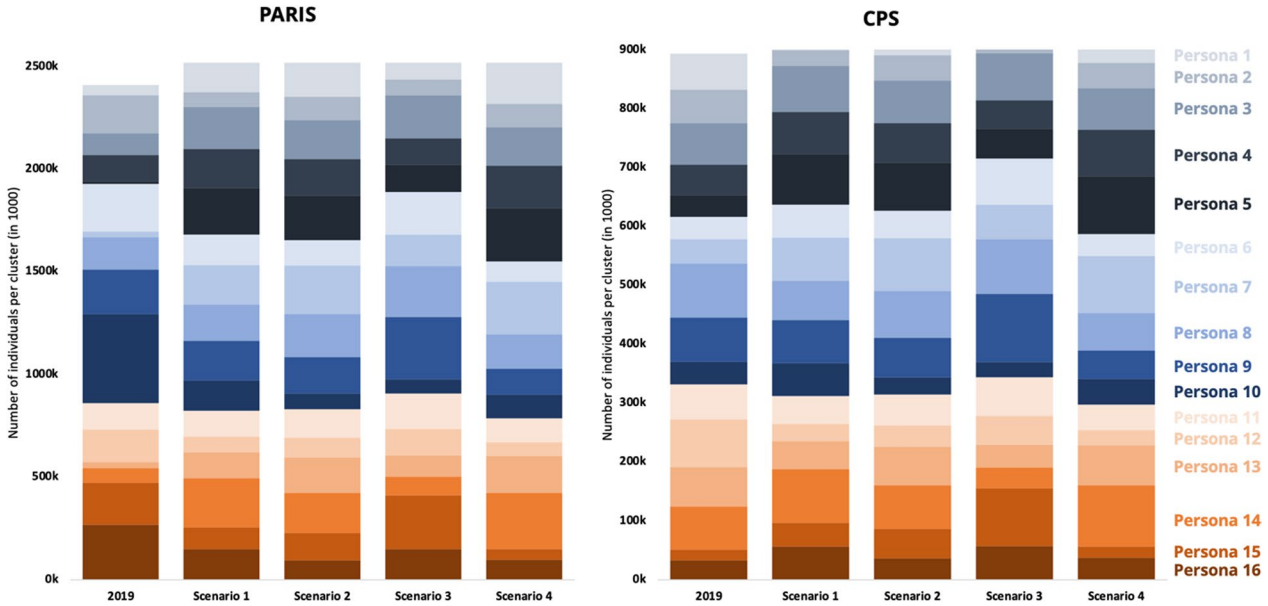

Fig. 6 Resulting distribution of PP (in 1000 s) for 2019 (census base year) and 2030 scenarios using outputs from reweighting and population growth rates per area

to the reweighting process. Applying this technique and analysing simulation outcomes based on these future SPs has been conducted separately in parallel research.

4 Scenario-specific sets of personas and synthetic populations

The first output is a quantitative distribution of PPs across scenarios. This data set remains linked to the initial census data and can aid the process of scaling up and redistributing the census entries to result in new quantitative insights on varying possible futures, as well as feed quantitative design approaches. Figure 6 shows the distributions in Paris and CPS.[6] The distribution for the 2019 base and the four archetypical scenarios is provided. We can analyse two information from the figure. First, all four scenarios have a higher population, clearly visible for Paris and slightly for CPS. Second, each PP is represented by a colour and changes its occurrence from one scenario to another.

The information from the PPs, enriched with information from the mobility survey, can feed the creation process of data-driven qualitative personas. An example persona for the PP10 (Annex, Fig. 7) has been created by adapting an existing persona template [55].

However, some limitations prevail. This paper focuses on the individual. Some data exists on household level and could find integration in the process. Inherent to working with futures, subjective inputs were used. While this can be supported via literature and expert validation, accuracy remains an uncertainty. The authors argue that the chance to be closer to the actual future is higher when working with several futures compared to working with a singular one or simply use the assumptions of present users and population as done predominantly in previous works.

5 Discussion and conclusion

This paper asked *how diverse future user needs can be integrated into design processes for urban mobility systems*. Future scenarios and different ways of modelling users were introduced that can support service design and simulation. Next, a mixed-method approach was proposed using existing qualitative (i.e., persona creation, scenario planning) and quantitative (i.e., k-proto clustering and agent-based simulation) methods to inform today's design decisions of urban mobility systems. Using these methods permits to (1) address pre-existing weaknesses (e.g., dominating qualitative approaches for persona creation, lack of user uncertainty integration in simulation), (2) build a basis for interdisciplinary collaboration between designers and

[6] The fitting process resulted in deviation values of maximum 3% between targeted values and actual results (Annex, Table 13).

data scientists, and (3) propose new components that can be used to adapt personas and SPs in a structured manner to future scenarios.

The method was used to generate validated data-driven proto-personas and reweigh them across future scenarios to allow the integration of trends and uncertainties in qualitative design and quantitative simulation. This results in a set of personas and SPs per scenario that enables the use of established people-centred design processes (i.e., personas) in a future context. In parallel, the proposed method results in a set of SPs that feed agent-based simulations of future scenarios. Aside from its field-specific contributions, the combined approach enables designers and data scientists to collaborate via shared and aligned artefacts.

5.1 Discussion

Existing works discussed in this article used extensively personas, SPs, and scenarios. Conceptualised by Grudin and Pruitt [27], *personas* have found ample application in mobility service design (e.g., [32, 33]. Vallet et al. [30] introduced future personas to consider future mobility user needs, while Al Maghraoui et al. [45] proposed methods to integrate future user experiences. While qualitative approaches dominate persona creation, Stevenson and Mattson [31] presented a quantitative method to generate personas.

SPs have found significant attention in mobility simulation and design, including replicable generation thereof [42] and various use cases in agent-based simulation (e.g., [43, 44]. Kamel et al. (2019) and Vosooghi et al. (2019) set a foundation for quantitative integration of diverse users, while Vallet et al. [48] proposed a conceptual framing of personas and SPs to advance people-centred urban mobility system design.

While each can play a crucial role in designing sustainable and people-centred urban mobility futures, the use of the analysed approaches is limited to specific fields and so far without consideration of multiple futures. Working with such scenarios can be critical to addressing rising uncertainty originating, among others, from the pace of technological advances, the climate crisis, and behavioural evolutions.

This paper proposes a method that uses these already established concepts to advance the work of designing with a people-centred perspective for multiple futures while permitting the collaboration of experts from different disciplines. It provides a novel generation approach for data-driven and realistic personas of today and tomorrow in the urban mobility context and responds to the lack of adapted future simulations. The outputs of this method have been used successfully to generate alternative SPs, test potential mobility solutions across future scenarios, and analyse impacts across PP groups in different geographical contexts.

Aside from its contributions, a range of limitations persist: Most notably the qualitative approach for defining future trends and uncertainties parameters. However, this challenge is inherent to working with futures and can be partially mitigated by basing assumptions on expert knowledge and existing research. Otherwise, a set of methodological choices has been made throughout the process. While each was made to respond to the identified challenges and needs, other methods exist which might be more adequate and should be tested as well.

5.2 Conclusion

In conclusion, the wicked problems of planning sustainable and people-centred urban mobility solutions today require a detailed consideration of the future impacts of socio-demographic and technological changes on solutions, users, and associated needs. The presented methodology comes with limitations inherent to its ambition and the underlying concepts. However, the authors argue that it makes a relevant contribution to people-centred urban mobility solution design and simulation processes, as well as creating artefacts for structured exchange between experts from qualitative and quantitative disciplines. The existing uptake across different use cases with different profiles of collaborators so far has shown its potential utility. The continuous extension and improvement, as well as wider testing and validation are nevertheless seen as crucial. This publication, supported by the use of open data and provided open-source code shall enable this process and ultimately permit designing more sustainable and people-centred solutions for the urban populations and mobility users of tomorrow.

Annex

See Tables 8, 9, 10, 11, 12, and 13; Fig. 7.

Table 8 Census information chosen for clustering (translated by first author)

I/H	Code	Description	Type	Use
Characteristics for categorisation				
I	AGED	Detailed age	Numerical	Category
H	AREA	Primary residence area	Categorical	Category
Household values				
H	NBPI	Number of rooms in the accommodation	Numerical	Cluster
H	NE5FR	Number of children aged 5 or under in the family (combined)	Numerical	Cluster
H	*NE17FR*	*Number of children aged 17 or under in the family (combined)*	*Numerical*	*Reference*
H	NPERR	Number of people in household	Numerical	Cluster
H	GARL	Reserved parking space	Categorical	Cluster
H	NA17	Economic activity in 17 positions	Categorical	Cluster
H	STOCD	Detailed housing occupancy status	Categorical	Cluster
H	TRANS	Main mode of transportation most often used to get to work	Categorical	Cluster
H	VOIT	Number of household cars	Numerical	Cluster
H	TYPL	Type of construction	Categorical	Cluster
H	SURF	Living area of residence	Ordinal	Cluster
Individual values				
I	COUPLE	Declaration of life as a couple	Dichotomous	Cluster
I	CS1	Socio-professional category in 8 positions	Categorical	Cluster
I	DIPL	Highest degree	Categorical	Cluster
I	ETUD	Registration in an educational institution	Dichotomous	Cluster
I	ILETUD	Study location	Categorical	Cluster
I	ILT	Work location	Categorical	Cluster
I	IMMI	Immigration status	Dichotomous	Cluster
I	INAI	Birthplace	Categorical	Cluster
I	MOCO	Way of living together	Categorical	Cluster
I	MODV	Way of life	Categorical	Cluster
I	SEXE	Gender	Dichotomous	Cluster
I	STAT_CONJ	Marital status	Categorical	Cluster
I	STATR	Grouped professional status	Categorical	Cluster
I	TACT	Activity type	Categorical	Cluster
I	TP	Working time	Categorical	Cluster

Table 9 Codes of used variables. Source: [50] Translated by first author

Code	Val.	Variable
GARL	1	With parking space(s) (FR: Avec emplacement(s) de stationnement)
	2	Without parking space (FR: Sans emplacement de stationnement)
	Y	Except principal residence (FR: Hors résidence principale)
	Z	Except regular accommodation (FR: Hors logement ordinaire)
NA17	AZ	Agriculture, forestry and fishing (FR: Agriculture, sylviculture et pêche)
	C1	Manufacture of foodstuffs, beverages and tobacco products (FR: Fabrication de denrées alimentaires, de boissons et de produits à base de tabac)
	C2	Coking and refining (FR: Cokéfaction et raffinage)
	C3	Manufacture of electrical, electronic and computer equipment; machinery manufacturing (FR: Fabrication d'équipements électriques, électroniques, informatiques; fabrication de machines)
	C4	Manufacture of transport equipment (FR: Fabrication de matériels de transport)

Table 9 (continued)

Code	Val.	Variable
	C5	Manufacture of other industrial products (FR: Fabrication d'autres produits industriels)
	DE	Extractive industries, energy, water, waste management and depollution (FR: Industries extractives, énergie, eau, gestion des déchets et dépollution)
	FZ	Construction
	GZ	Trade; automobile and motorcycle repair (FR: Commerce; réparation d'automobiles et de motocycles)
	HZ	Transport and storage (FR: Transports et entreposage)
	IZ	Accommodation and catering (FR: Hébergement et restauration)
	JZ	Information and/et communication
	KZ	Financial and insurance activities (FR: Activités financières et d'assurance)
	LZ	Real estate activities (FR: Activités immobilières)
	MN	Scientific and technical activities; administrative and support services (FR: Activités scientifiques et techniques; services administratifs et de soutien)
	OQ	Public administration, education, human health and social action (FR: Administration publique, enseignement, santé humaine et action sociale)
	RU	Other service activities (FR: Autres activités de services)
	ZZ	Not applicable (FR: Sans objet)
STOCD	00	Ordinary unoccupied dwelling (FR: Logement ordinaire inoccupé)
	10	Owner (FR: Propriétaire)
	21	Tenant or sub-tenant of an empty non-HLM rented accommodation (FR: Locataire ou sous-locataire d'un logement loué vide non HLM)
	22	Tenant or sub-tenant of empty rented housing HLM (FR: Locataire ou sous-locataire d'un logement loué vide HLM)
	23	Tenant or sub-tenant of furnished rented accommodation or a hotel room (FR: Locataire ou sous-locataire d'un logement loué meublé ou d'une chambre d'hôtel)
	30	Free accommodation (FR: Logé gratuitement)
	ZZ	Except ordinary housing (FR: Hors logement ordinaire)
TRANS	1	No transportation (FR: Pas de transport)
	2	Walking (or rollerblading, scooter) (FR: Marche à pied (ou rollers, patinette))
	3	Bicycle (including electrically assisted) (FR: Vélo (y compris à assistance électrique))
	4	Motorised two-wheelers (FR: Deux-roues motorisé)
	5	Car, truck, van (FR: Voiture, camion, fourgonnette)
	6	Public transport (FR: Transports en commun)
	Z	Not applicable (FR: Sans objet)
TYPL	1	House (FR: Maison)
	2	Apartment (FR: Appartement)
	3	Residential accommodation (FR: Logement-foyer)
	4	Hotel room (FR: Chambre d'hôtel)
	5	Makeshift dwelling (FR: Habitation de fortune)
	6	Independent room (having its own entrance) (FR: Pièce indépendante (ayant sa propre entrée))
	Z	Excluding ordinary housing (FR: Hors logement ordinaire)
SURF	1	Less than 30 sqm (FR: Moins de 30 m2)
	2	Between 30 and 40 sqm (FR: De 30 à moins de 40 m2)
	3	Between 40 and 60 sqm (FR: De 40 à moins de 60 m2)
	4	Between 60 and 80 sqm (FR: De 60 à moins de 80 m2)
	5	Between 80 and 100 sqm (FR: De 80 à moins de 100 m2)
	6	Between 100 and 120 sqm (FR: De 100 à moins de 120 m2)
	7	120 sqm or more (FR: 120 m2 ou plus)
	Y	Except main residence (FR: Hors résidence principale)
	Z	Except regular accommodation (FR: Hors logement ordinaire)
Couple	1	Declared to live with a partner (FR: A déclaré vivre en couple)
	2	Declared to not live with a partner (FR: A déclaré ne pas vivre en couple)
CS1	1	Farmes (FR: Agriculteurs exploitants)
	2	Craftsmen, merchants and entrepreneurs (FR: Artisans, commerçants et chefs d'entreprise)
	3	Executives and higher intellectual professions (FR: Cadres et professions intellectuelles supérieures)
	4	Intermediate professions (FR: Professions Intermédiaires)
	5	Employees (FR: Employés)

Table 9 (continued)

Code	Val.	Variable
	6	Workers (FR: Ouvriers)
	7	Pensioner (FR: Retraités)
	8	Other people without professional activity (FR: Autres personnes sans activité professionnelle)
DIPL	01	No schooling or stopping before the end of primary school (FR: Pas de scolarité ou arrêt avant la fin du primaire)
	02	No diploma and schooling interrupted at the end of primary school or before the end of college (FR: Aucun diplôme et scolarité interrompue à la fin du primaire ou avant la fin du college)
	03	No degree and schooling until the end of college or beyond (FR: Aucun diplôme et scolarité jusqu'à la fin du collège ou au-delà)
	11	CEP (certificate of primary studies) (FR: CEP (certificat d'études primaires))
	12	BEPC, elementary patent, college certificate, DNB (FR: BEPC, brevet élémentaire, brevet des collèges, DNB)
	13	CAP, BEP or diploma of equivalent level (FR: CAP, BEP ou diplôme de niveau équivalent)
	14	General or technological baccalaureate, higher patent, capacity in law, DAEU, ESEU (FR: Baccalauréat général ou technologique, brevet supérieur, capacité en droit, DAEU, ESEU)
	15	Vocational baccalaureate, professional, technician or teaching certificate, equivalent diploma (FR: Baccalauréat professionnel, brevet professionnel, de technicien ou d'enseignement, diplôme équivalent)
	16	BTS, DUT, Deug, Deust, bac + 2 level health or social diploma, equivalent diploma (FR: BTS, DUT, Deug, Deust, diplôme de la santé ou du social de niveau bac + 2, diplôme équivalent)
	17	Bachelor's degree, pro license, master's degree, equivalent diploma at bac + 3 or bac + 4 level (FR: Licence, licence pro, maîtrise, diplôme équivalent de niveau bac + 3 ou bac + 4)
	18	Master, DEA, DESS, high school diploma level bac + 5, doctorate in health (FR: Master, DEA, DESS, diplôme grande école niveau bac + 5, doctorat de santé)
	19	Research doctorate (excluding health) (FR: Doctorat de recherche (hors santé))
	ZZ	Out of scope (under 14) (FR: Hors champ (moins de 14 ans))
	YY	Except principal residence (FR: Hors résidence principale)
ETUD	1	Yes (FR: Oui)
	2	No (FR: Non)
ILETUD/ILT	1	In the current municipality of residence (FR: Dans la commune de résidence actuelle)
	2	In another municipality of the department of residence (FR: Dans une autre commune du département de résidence)
	3	In another department of the region of residence (FR: Dans un autre département de la région de résidence)
	4	Outside the region of current residence: in mainland France (FR: Hors de la région de résidence actuelle: en métropole)
	5	Outside the region of current residence: in a DOM (FR: Hors de la région de résidence actuelle: dans un DOM)
	6	Outside the region of current residence: in a COM (FR: Hors de la région de résidence actuelle: dans une COM)
	7	Abroad (FR: À l'étranger)
	Z	Not applicable (no registration in an educational institution) (FR: Sans objet (pas d'inscription dans un établissement d'enseignement))
IMMI	1	Immigrant (FR: Immigrés)
	2	Non-immigrant (FR: Non immigrés)
INAI	1	In the department of current residence (FR: Dans le département de résidence actuelle)
	2	In another department of the region of current residence (FR: Dans un autre département de la région de résidence actuelle)
	3	Outside the region of current residence: in mainland France (FR: Hors de la région de résidence actuelle: en métropole)
	4	Outside the region of current residence: in a DOM (FR: Hors de la région de résidence actuelle: dans un DOM)
	5	Outside the region of current residence: in a TOM-COM (FR: Hors de la région de résidence actuelle: dans un TOM-COM)
	6	Abroad (FR: À l'étranger)
MOCO	11	Children of a couple (FR: Enfants d'un couple)
	12	Children from a single parent family (FR: Enfants d'une famille monoparentale)
	21	Adults of a childless couple (FR: Adultes d'un couple sans enfant)
	22	Adults of a couple with child(ren) (FR: Adultes d'un couple avec enfant(s))
	23	Adults from a single-parent family (FR: Adultes d'une famille monoparentale)
	31	Non-family in multi-person household (FR: Hors famille dans ménage de plusieurs personnes)
	32	People living alone (FR: Personnes vivant seules)
	40	Persons living outside the household (FR: Personnes vivant hors ménage)
MODV	11	Children of a couple (FR: Enfants d'un couple)
	12	Children from a single parent family (FR: Enfants d'une famille monoparentale)

Table 9 (continued)

Code	Val.	Variable
	20	Single people under 40 years (FR: Personnes seules de moins de 40 ans)
	31	Members under 40 of a childless couple (FR: Membres de moins de 40 ans d'un couple sans enfant)
	32	Members of a couple with children (FR: Membres d'un couple avec enfants)
	33	Parents of a single-parent family (FR: Parents d'une famille monoparentale)
	40	Members aged 40 or over of a childless couple (FR: Membres de 40 ans ou plus d'un couple sans enfant)
	50	Single people aged 40 or over (FR: Personnes seules de 40 ans ou plus)
	60	Persons living outside the family in a multi-person household (FR: Personnes vivant hors famille dans un ménage de plusieurs personnes)
	70	Persons living outside the household (FR: Personnes vivant hors ménage)
SEXE	1	Men (FR: Hommes)
	2	Woman (FR: Femmes)
STAT_CONJ	1	Married (FR: Marié(e))
	2	In a registered partnership (FR: Pacsé(e))
	3	In concubinage or free union (FR: En concubinage ou union libre)
	4	Widow (FR: Veuf, Veuve)
	5	Divorced (FR: Divorcé(e))
	6	Single (FR: Célibataire)
STATR	1	Employees (FR: Salariés)
	2	Non salaried (FR: Non salariés)
	Z	Not applicable (FR: Sans objet)
TACT	11	Workers with a job, including apprenticeship or paid internship. (FR: Actifs ayant un emploi, y compris sous apprentissage ou en stage rémunéré.)
	12	Unemployed (FR: Chômeurs)
	21	Retired or pre-retired (FR: Retraités ou préretraités)
	22	Pupils, students, unpaid interns aged 14 or over (FR: Élèves, étudiants, stagiaires non rémunéré de 14 ans ou plus)
	23	Under 14 years (FR: Moins de 14 ans)
	24	Housewives or men (FR: Femmes ou hommes au foyer)
	25	Other inactive (FR: Autres inactifs)
TP	1	Full time (FR: Temps complet)
	2	Part time (FR: Temps partiel)
	Z	Not applicable (FR: Sans objet)

Table 10 Descriptive statistics for numeric variables

	Individual	Household level				
	AGED	NBPI	NE5FR	NE17FR	NPERR	VOIT
Min.	0	1	0	0	1	0
1st Qu.	20	3	0	0	2	0
Median	38	4	0	1.24	3	1
Mean.	39	3.76	0.41	1.24	3.09	1.07
3rd Qu.	56	5	0.43	2	4	2
Max.	120	20	4	4	6	3

Table 11 Descriptive statistics for categorical data after population scaling, part 1

SEXE			COUPLE	ETUD	IMMI
Male	1967 k	Yes	6004 k	3550 k	2719 k
Female	2119 k	No	7654 k	10,109 k	10,939 k

Table 12 Descriptive statistics for categorical data after population scaling, part 2

GARL		TP		STATR	
Parking	7905 k	Fulltime	5265 k	Employee	7526 k
No parking	5520 k	Parttime	868 k	Non-employee	5504 k
Distant	233 k	NA	7526 k	NA	628 k

Table 13 Initial, target, and actual proportional distributions of PP across scenarios

Scenario	Attribute	Value	Initial	Target	Result	Deviation (%)
1	Household	Mean	2.992	2.992	2.993	0
	Cars	Mean	1.057	1.110	1.105	0
	Location	1	0.174	0.188	0.185	1
	Location	2	0.173	0.185	0.188	−2
	Location	3	0.346	0.306	0.310	−1
	Location	nA	0.307	0.317	0.317	0
2	Household	Mean	2.992	2.842	2.844	0
	Cars	Mean	1.057	1.004	1.007	0
	Location	1	0.174	0.188	0.187	0
	Location	2	0.173	0.190	0.184	3
	Location	3	0.346	0.319	0.320	−1
	Location	nA	0.307	0.308	0.309	0
3	Household	Mean	2.992	3.291	3.282	0
	Cars	Mean	1.057	1.004	1.016	−1
	Location	1	0.174	0.188	0.189	−1
	Location	2	0.173	0.193	0.185	4
	Location	3	0.346	0.325	0.322	1
	Location	nA	0.307	0.303	0.303	0
4	Household	Mean	2.992	2.692	2.700	0
	Cars	Mean	1.057	1.057	1.050	1
	Location	1	0.174	0.188	0.186	1
	Location	2	0.173	0.182	0.184	−1
	Location	3	0.346	0.300	0.307	−2
	Location	nA	0.307	0.322	0.323	0

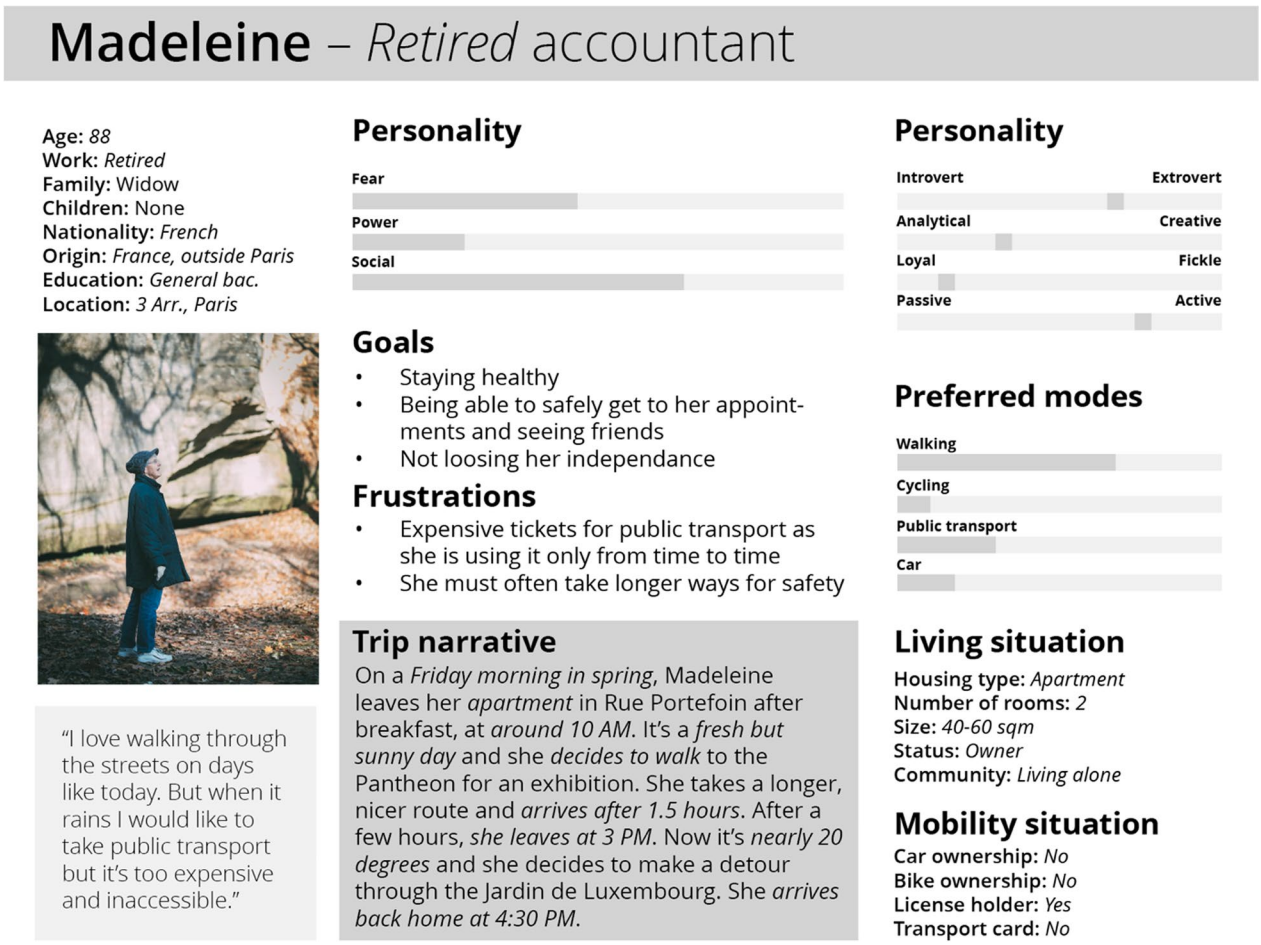

Fig. 7 Exemplary persona card for PP16. Everything written in italic is coming from underlying data. All other were deduced from it when and where possible. Persona format adapted from template by xtensio [55]

Acronyms

CPS	Inter-council partnership (FR: Communauté d'Agglomeration) Paris-Saclay
EMP	Mobility of people survey (FR: Enquête Mobilité des personnes)
GHG	Greenhouse Gas
IDF	Île-de-France
INSEE	National Institute of Statistics and Economic Studies (FR: Institut national de la statistique et des études économiques)
IRIS	Units Grouped for Statistical Information (FR: Ilots Regroupés pour l'Information Statistique)
MGP	Metropole du Grand Paris
PP	Proto-Persona
SDG	Sustainable Development Goal
SP	Synthetic Population

Acknowledgements
This research work has been carried out at the Anthropolis Chair and we thank our colleagues and the anonymous reviewers for the many helpful comments.

Authors' contribution
TG, FV, and SH conceived, designed, and executed the study. TG drafted the manuscript. BY contributed to conceptualisation, literature review and played a key role in manuscript review and editing. All authors have read and approved the final version of the manuscript.

Funding
This work has been supported by the French government under the 'France 2030' program, as part of the SystemX Technological Research Institute.

Availability of data and materials
Used data is open access and available via the information provided in the references. Data files of the output as well as all files used for the data analysis and transformation are available as commented files online: https://github.com/TjarkGall/proto-persona-clustering.

Declarations

Competing interests
All authors declares that they have no competing interests.

Received: 1 November 2022 Accepted: 14 November 2023
Published online: 28 November 2023

References

1. IPCC (2021) Summary for policymakers. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth

Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press.
2. CDC (2020) Road traffic injuries and deaths: A global problem. Accessible at: https://www.cdc.gov. Last accessed 03 Dec 2020.
3. Cebr. (2014). The future economic and environmental costs of gridlock in 2030. An assessment of the direct and indirect economic and environmental costs of idling in road traffic congestion to households in the UK, France, Germany and the USA. Centre for Economics and Business Research (Cebr).
4. Climate Watch. (2020). Historical GHG Emissions. Accessible at: https://www.climatewatchdata.org [last accessed 28 October 2022].
5. Metabolic (2019) *Metal demand for electric vehicle: Recommendations for fair, resilient, and circular transport systems*. Amsterdam: Metabolic.
6. UN. (2015). Draft outcome document of the United Nations summit for the adoption of the post-2015 development agenda. Transforming our world: The 2030 Agenda for Sustainable Development.
7. Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy, 15*(2), 73–80. https://doi.org/10.1016/j.tranpol.2007.10.005
8. Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155–169.
9. Hörl, S., & Balac, M. (2021). Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies, 130*, 103291. https://doi.org/10.1016/j.trc.2021.103291
10. Gall, T., Vallet, F., Douzou, S., Yannou, B. (2021) Re-defining the system boundaries of human-centred design. *Proceedings of the Design Society, 1*, 2521–2530. https://doi.org/10.1017/pds.2021.513
11. Spaniol, M. J., & Rowland, N. J. (2019). Defining scenario. Futures Foresight. *Science, 1*, e1. https://doi.org/10.1002/ffo2.3
12. Börjeson, L., Höjer, M., Dreborg, K.-H., Ekvall, T., & Finnveden, G. (2006). Scenario types and techniques: Towards a user's guide. *Futures, 38*(7), 723–739. https://doi.org/10.1016/j.futures.2005.12.002
13. Goodspeed, R. (2020). *Scenario planning for cities and regions: Managing and Envisioning Uncertain Futures*, Washington: Lincoln Institute of Land Policy.
14. Banister, D., & Hickman, R. (2013). Transport futures: Thinking the unthinkable. *Transport Policy, 29*, 283–193. https://doi.org/10.1016/j.tranpol.2012.07.005
15. Miskolczi, M., Földes, D., Munkácsy, A., Jászberényi (2021). Urban mobility scenarios until the 2030s. *Sustainable Cities and Society, 72*, 103029. https://doi.org/10.1016/j.scs.2021.103029
16. Pucci, P. (2021). Spatial dimensions of electric mobility—Scenarios for efficient and fair diffusion of electric vehicles in the Milan Urban Region, Cities, 110, 103069. https://doi.org/10.1016/j.cities.2020.103069/
17. Sharma, I., Padmanabhi, R., Dikshit, A. K., Chandel, M. K. (2023). Urban transport emissions under current and alternative mitigation policy scenarios for the Mumbai Metropolitan region. *Case Studies on Transport Policy*, 101001. https://doi.org/10.1016/j.cstp.2023.101001
18. Soria-Lara, J. A., Ariza-Álvarez, A., Aguilera-Benavente, F., Cascajo, R, Arce-Ruiz, R. M., López, C., Gómez-Delgado, M. (2021) Participatory visioning for building disruptive future scenarios for transport and land use planning. *Journal of Transport Geography, 90*, 102907. https://doi.org/10.1016/j.jtrangeo.2020.102907
19. Dator, J. (2019). What futures studies is, and is not. In Jim Dator: *A noticer in time. Anticipation Science*, vol. 5, Cham: Springer. https://doi.org/10.1007/978-3-030-17387-6_1
20. Fergnani, A., & Song, Z. (2020). The six scenario archetypes framework: A systematic investigation of science fiction films set in the future. *Futures, 124*, 102645. https://doi.org/10.1016/j.futures.2020.102645
21. Shaaban, M., Voglhuber-Slavinsky, A., Dönitz, E., Macpherson, J., Paul, C., Mouratiadou, I., Helming, K., Piorr, A. (2023). Understanding the future and evolution of agri-food systems: A combination of qualitative scenarios with agent-based modelling. *Futures, 149*, 103141. https://doi.org/10.1016/j.futures.2023.103141.
22. Schlenther, T., Wagner, P., Rybczak, G., Nagel, K., Bieker-Walz, L., Ortgiese, M. (2022). Simulation-based investigation of transport scenarios for Hamburg. *Procedia Computer Science, 201*, 587–593. https://doi.org/10.1016/j.procs.2022.03.076
23. Maheshwari, T., Fourie, P., Ordoñez Medina, S. A., & Axhausen, K. W. (2023). Iterative urban design and transport simulation using Sketch MATSim. *Journal of Urban Design*. https://doi.org/10.1080/13574809.2023.2214080
24. Cooper, A. (1999). *The inmates are running the asylum*. Macmillan.
25. Bornet, C., & Brangier, É. (2013). La méthode des personas : Principes, intérêts et limites. *Bulletin de Psychologie, 2*(524), 115–134. https://doi.org/10.3917/bupsy.524.0115
26. Adlin, T., & Pruitt, J. (2010). *The essential persona lifecycle: Your guide to building and using personas*. Morgan Kaufmann/Elsevier.
27. Grudin, J., & Pruitt, J. (2002). Personas, participatory design and product development: An infrastructure for engagement. In *Proceedings of participation and design conference (PDC2002)*, Sweden, pp. 144–161.
28. Salminen, J., et al. (2020). Persona transparency: Analyzing the impact of explanations on perceptions of data-driven personas. *International Journal of Human-Computer Interaction, 36*(8), 788–800. https://doi.org/10.1080/10447318.2019.1688946/
29. Goodman-Deane, J., Bradley, M., Waller, S., & Clarkson, P. (2021). Developing personas to help designers to understand digital exclusion. *Proceedings of the Design Society, 1*, 1203–1212. https://doi.org/10.1017/pds.2021.120/
30. Vallet, F., Puchinger, J., Millonig, A., Lamé, G., & Nicolaï, I. (2020) Tangible futures: Combining scenario thinking and personas—a pilot study on urban mobility. *Futures, 2020*, 117. https://doi.org/10.1016/j.futures.2020.102513
31. Stevenson, P. D., & Mattson, C. A. (2019). The personification of big data. In *International conference on engineering design ICED19*, Delft, August 2019. https://doi.org/10.1017/dsi.2019.409
32. Kong, P., Cornet, H., Frenkler, F. (2018). Personas and emotional design for public service robots: A case study with autonomous vehicles in public transportation. 2018 International Conference on Cyberworlds (CW), pp. 284–287. https://doi.org/10.1109/CW.2018.00058
33. Gargiulo, E., Giannantonio, R., Guercio, E., Borean, C., & Zenezini, G. (2015). Dynamic ride sharing service: Are users ready to adopt it? *Procedia Manufacturing, 3*, 777–784. https://doi.org/10.1016/j.promfg.2015.07.329
34. Hermes, K., & Poulsen, M. (2012). A review of current methods to generate synthetic spatial microdata using reweighting and future directions. *Computers, Environment and Urban Systems, 36*, 281–290. https://doi.org/10.1016/j.compenvurbsys.2012.03.005
35. Ramadan, O. E., & Sisiopiku, V. P. (2020) A critical review on population synthesis for activity- and agent-based transportation models. In: De Luca, S., Di Pace, R., Djordjevic, B. (Eds.), *Transportation systems analysis and assessment*. IntechOpen. https://doi.org/10.5772/intechopen.86307
36. Durán-Heras, A., García-Gutiérrez, I., Castilla-Alcalá, G. (2018) Comparison of iterative proportional fitting and simulated annealing as synthetic population generation techniques: Importance of the rounding method. *Computers, Environment and Urban Systems 68*, 78–88. https://doi.org/10.1016/j.compenvurbsys.2017.11.001
37. Yameogo, B. F., Vandanjon, P.-O., Gastineau, P., Hankach, P. (2021). Generating a two-layered synthetic population for French municipalities: Results and evaluation of four synthetic reconstruction methods. *JASSS-Journal of Artificial Societies and Social Simulation 24*, 5. https://doi.org/10.18564/jasss.4482
38. Sun, L., & Erath, A. (2015). A Bayesian network approach for population synthesis. *Transportation Research Part C: Emerging Technologies, 61*, 49–62. https://doi.org/10.1016/j.trc.2015.10.010
39. Saadi, I., Mustafa, A., Teller, J., Farooq, B., & Cools, M. (2016). Hidden Markov model-based population synthesis. *Transportation Research Part B: Methodological, 90*, 1–21. https://doi.org/10.1016/j.trb.2016.04.007
40. Borysov, S. S., Rich, J., & Pereira, F. C. (2019). How to generate micro-agents? A deep generative modeling approach to population synthesis. *Transportation Research Part C: Emerging Technologies, 106*, 73–97. https://doi.org/10.1016/j.trc.2019.07.006
41. Saadi, I., Farooq, B., Mustafa, A., Teller, J., & Cools, M. (2018). An efficient hierarchical model for multi-source information fusion. *Expert Systems with Applications, 110*, 352–362. https://doi.org/10.1016/j.eswa.2018.06.018
42. Hörl, S., Balac, M., & Axhausen, K. W. (2019) Dynamic demand estimation for an AMoD system in Paris. In *2019 IEEE intelligent vehicles symposium (IV). Presented at the 2019 IEEE intelligent vehicles symposium (IV)*, IEEE, Paris, France, pp. 260–266. https://doi.org/10.1109/IVS.2019.8814051

43. Le Bescond, V., Can, A., Aumond, P., & Gastineau, P. (2021). Open-source modelling chain for the dynamic assessment of road traffic noise exposure. *Transportation Research Part D: Transport and Environment, 94*, 102793. https://doi.org/10.1016/j.trd.2021.102793
44. Panos, E., & Margelou, S. (2019). Long-term solar photovoltaics penetration in single- and two-family house in Switzerland. *Energies, 12*, 2460. https://doi.org/10.3390/en12132460
45. Al Maghraoui, O., Vallet, F., Puchinger, J., & Yannou, B. (2019). Modeling traveler experience for designing urban mobility systems. *Design Science, 5*, E7. https://doi.org/10.1017/dsj.2019.6/
46. Kamel, J., Vosooghi, R., Puchinger, J., Ksontini, F., & Sirin, G. (2019). Exploring the impact of user preferences on shared autonomous vehicle modal split: A multi-agent simulation approach. *Transportation Research Procedia, 37*, 115–122. https://doi.org/10.1016/j.trpro.2018.12.173
47. Vosooghi, R., Kamel, J., Puchinger, J. Leblond, V. & Jankovic, M. (2019) Robo-Taxi service fleet sizing: assessing the impact of user trust and willingness-to-use. *Transportation*, *46*. https://doi.org/10.1007/s11116-019-10013-x
48. Vallet, F., Hörl, S., & Gall, T. (2022). Matching synthetic populations with personas: A test application for urban mobility. *Proceedings of the Design Society, 2*, 1795–1804. https://doi.org/10.1017/pds.2022.182/
49. INSEE (2022) Recensement de la population. Individus localisés au canton-ou-ville en 2019. Accessible at: https://www.insee.fr/fr/statistiques/6544333. Last accessed 2 July 2022.
50. EMP. (2021). Résultats détaillés de l'enquête mobilité des personnes de 2019. Ministère de la Transition Écologique et de la Cohésion des Territoires. Accessible at: https://www.statistiques.developpement-durable.gouv.fr/resultats-detailles-de-lenquete-mobilite-des-personnes-de-2019. Last accessed 23 Oct 2022.
51. Gall, T., Vallet, F., Yannou, B. (2023) Comment concevoir des systèmes de mobilité urbaine pour les citadins du futur ? 12e Colloque EPIQUE, Paris, 5–7 July 2023.
52. Szepannek, G. (2018) clustMixType: User-friendly clustering of mixed-type data in R. *The R Journal, 10*, 2. ISSN 2073-4859.
53. INSEE. (2017). Même vieillissante, l'Île-de-France resterait la région la plus jeune de France métropolitaine en 2050. Accessible at: https://www.insee.fr/fr/statistiques/2866751. Last accessed 24 Oct 2022.
54. Fergnani, A., & Jackson, M. (2019). Extracting scenario archetypes: A quantitative text analysis of documents about the future. *Future & Foresight Science*, *½*, 1–14. https://doi.org/10.1002/ffo2.17
55. Xtensio. (2022). Persona template. Accessible at: https://app.xtensio.com/design/aqk17wkn. Accessed 28 Oct 2022.
56. INSEE (2011) La population active en métropole à l'horizon 2030 : une croissance significative dans dix régions. Accessible at: https://www.insee.fr/fr/statistiques/2124013. Last accessed 24 Oct 2022.
57. INSEE. (2021). En 2017, les ménages consacrent 11% de leur revenu disponible à la voiture. Accessible at: https://www.insee.fr/fr/statistiques/5358250. Last accessed 28 Oct 2022.
58. L'Institut Paris Region. (2020). *Quel role pour le mass transit en Île-de-France à l'heure de la crise sanitaire ?* Note Rapid, no. 864. September 2020.
59. OECD. (2011). The future of families: A synthesis report. Paris: OECD.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com