Cross-domain transfer in mobility-based socio-demographic prediction: Evidence from French and Swiss mobility datasets

Eliane Casassa*[1,2], Etienne Côme[2], Renaud Oger[1], Kay Axhausen[3], and Latifa Oukhellou[2]

[1]MOP/ETG, Régie Autonome des Transports Parisiens (RATP), Fontenay-sous-Bois, France
[2]Université Gustave Eiffel, COSYS-GRETTIA, Champs-sur-Marne, France
[3]Institute for Transport Planning and Systems, ETH Zürich, Switzerland

Short summary

The widespread adoption of geolocation technologies has transformed the analysis of human mobility, offering a scalable alternative to traditional household travel surveys (HTS). However, such datasets typically lack socio-demographic labels, making it challenging to assess their representativeness. In this paper, we address this issue through an indirect evaluation strategy based on cross-domain model transfer. Models trained to infer socio-demographic attributes from mobility trajectories are applied to an unlabeled target dataset, where they are used to reconstruct aggregate population distributions. Using French and Swiss mobility datasets, we evaluated the robustness of this transfer approach. Our results show that dataset shifts significantly limit model portability, leading to substantial degradation in individual-level predictive performance. We then investigate quantification as a post hoc correction for the prior shift and demonstrate that Probabilistic Classify and Count (PCC) remains effective in estimating class prevalence under distributional changes. Finally, we highlight promising directions for improving cross-domain generalization, including the learning of domain-invariant representations and the development of alignment strategies to address covariate and concept shifts.
Keywords: cross-domain transfer, mobility, quantification, socio-demographic inference, tracking-based travel survey, transfer learning

1 Introduction

The widespread adoption of geolocation technologies has transformed human mobility analysis. Human activities now generate continuous digital traces, particularly spatio-temporal trajectories, that capture individual movements with unprecedented granularity (González et al., 2008). Previous research has demonstrated that individual mobility trajectories carry powerful implicit signals about socioeconomic status, age, or gender (Almaatouq et al., 2016; Wu et al., 2019; Bakhtiari et al., 2023).

Despite their implicit richness, raw trajectory datasets (such as smart card data or other passive data sources) often lack associated socio-demographic labels. This absence limits their use in applications that require statistical representativeness. Traditionally, this information is captured by Household Travel Surveys (HTS), which provide high ground-truth detail, but are costly and infrequent (Stopher & Greaves, 2007). This gap motivates the use of inference methods to enrich raw mobility datasets. In this context, two objectives are pursued: either inferring socio-economic attributes at the individual level to enrich user profiles or estimating the global class proportions to evaluate and account for collection biases in further analyzes.

These objectives fall within the scope of transfer learning, which examines how knowledge can be transferred between a source domain (labeled training data) and a target domain (unlabeled deployment data). In real-world mobility studies, this often implies a cross-domain setting, where models must generalize across different geographical areas or collection protocols.

To address these objectives, our study investigates the transferability of socio-demographic inference models between two distinct mobility datasets: a French GPS-based travel survey conducted in the Paris region (EMG) and a Swiss smartphone-based tracking survey (TimeUse+).

In cross-domain transfer, models are challenged by the phenomenon of dataset shift (Quiñonero-Candela et al., 2009), where the joint distribution of features and attributes differs between the

source and target domains ($P_S(X,Y) \neq P_T(X,Y)$). In this formulation, X denotes the feature space of mobility descriptors derived from individual spatio-temporal trajectories, Y corresponds to the discrete socio-demographic attributes to be inferred, and the subscripts S and T are used to distinguish between the source domain (France) and target domain (Swiss), respectively. As illustrated in figure 1, such a shift is generally categorized into three fundamental types, which may occur independently or simultaneously Moreno-Torres et al. (2012):

- *A Prior shift* occurs when the proportions of the classes differ between the domains, i.e. when $P_S(Y) \neq P_T(Y)$. In our context, this reflects the structural differences in the socio-demographic composition between the source (France) and target (Switzerland) study groups.
- *A Covariate shift* arises when the marginal distribution of features changes, i.e. when $P_S(X) \neq P_T(X)$. This reflects variations in travel patterns.
- *A Concept shift* happens when the posterior distribution itself changes, $P_S(Y|X) \neq P_T(Y|X)$. This implies that the decision boundary differs between the two datasets.

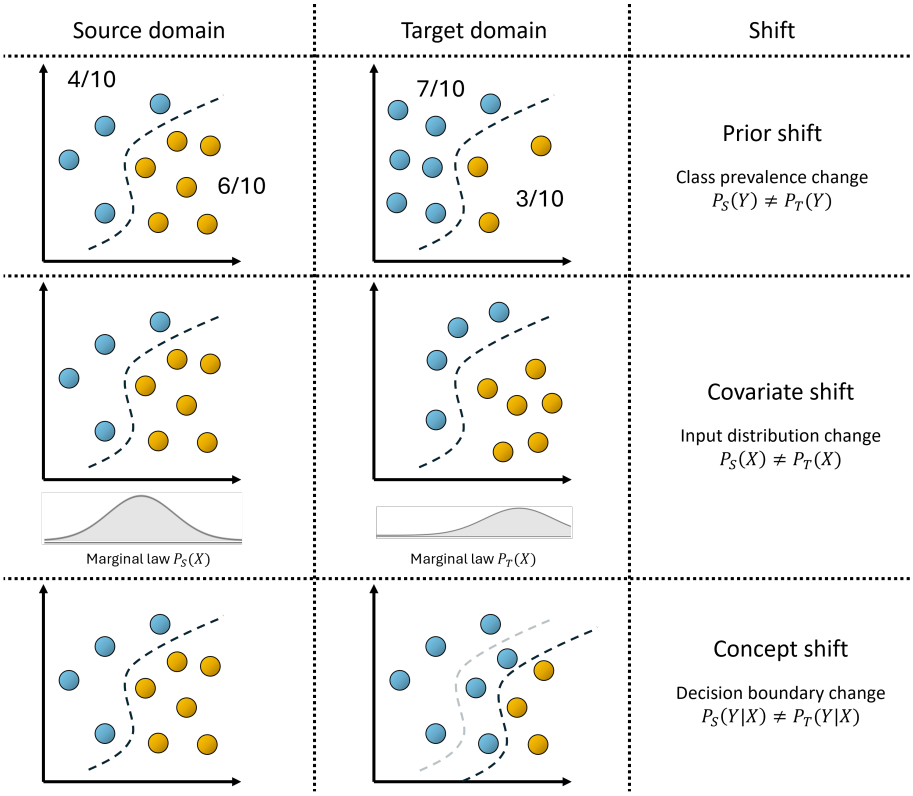


Figure 1: Illustration of dataset shift types for a binary classification task, when a model is trained on a source domain D_S different from the target domain D_S. Each point represents an individual. Individual mobility is represented by spatial coordinates, and class membership by colors. Decision boundary is represented in dotted line.

Recognizing these shifts as a major barrier to cross-domain transfer, we explore strategies to mitigate their impact. Although individual-level inference is valuable, a primary objective of this study is to recover global class proportions in the target domain. To this end, we specifically focus on quantification methods. Formally, quantification is the task of estimating the class prevalence $P(Y)$ of a target population, given a labeled source domain. These methods are specifically designed to address prior shift but operate under the assumption that no concept shift occurs. A wide range of quantification methods have been proposed, for a comprehensive comparative evaluation of their performance, we refer the reader to Schumacher et al. (2025).
In light of these challenges, the use of the EMG (France) and TimeUse+ (Switzerland) datasets provides a robust comparative framework to evaluate how dataset shifts affect model robustness across heterogeneous domains. Furthermore, this setup serves to identify the potential of quantification-based corrective methods to redress the global distribution $P(Y)$. To guide this investigation, we formulate the following research questions:

- When applying the model trained on French data to the Swiss context, to what extent does the dataset shift degrade individual-level inference performance?
- What is the nature of the underlying data shifts that account for this performance decline?
- To what extent can quantification algorithms serve as a valid post-processing corrective tool to adjust the bias in global distribution estimates?

The remainder of this paper is organized as follows. Section 2 describes the datasets, the feature alignment pipeline, and the methodology for both inference and quantification. Section 3 presents the experimental results, from intra-domain benchmarks to cross-domain transfer analysis and the evaluation of quantification methods. Finally, Section 4 summarizes our findings and outlines directions for future research.

2 Methodology

Data collection and feature alignments

This study uses two high-resolution mobility datasets collected between 2022 and 2023: the French EMG and the Swiss TimeUse+ (Meister et al., 2020; Winkler et al., 2024). The EMG dataset tracked 3,300 participants (aged 16–80) in the Paris region for seven days using GPS loggers (2s sampling frequency), validated by daily diary and interviews. Conversely, TimeUse+ passively recorded 1,318 participants (> 18 years) in German-speaking Switzerland over four weeks via a smartphone app (MotionTag SDK), with digital diary validations.
To align observation windows for cross-domain analysis, we extracted a single week of data per TimeUse+ participant to match the EMG duration. We randomly selected a unique week starting on a Monday from periods containing at least four validated days.
Socio-demographic variables were harmonized into a shared categorical space for four target attributes: gender, presence of dependent children, age (< 25, 25–44, 45–64 and > 65 years), and diploma level. To account for different educational systems, diploma was simplified into a binary attribute (university degree vs. under university). The diploma variable serves as a proxy for socioeconomic status, as income data was unfortunately unavailable in the French EMG dataset. However, the correlation between degrees and income levels appears to differ culturally. Swiss professional contexts often prioritize experience, whereas the French system remains aligned with academic diplomas.
Each participant is thus represented by their attributes and a weekly mobility sequence (trips, times, distances, speeds, coordinates). Following our previous work (Casassa et al., 2026), these raw sequences are transformed into aggregated individual-level feature vectors of 68 descriptors (table 1).

Table 1: Descriptors used for inference model.

Category	**Descriptors**	**Size**
Temporal features	Mean trip duration on weekdays / weekends	2
	Earliest and latest departure times during the survey period	2
	Number of short trips per day (< 15 minutes)	1
	Number of long trips per day (> 90 minutes)	1
Spatial dispersion	Radius of gyration (weighted by time spent in each place)	1
	Area of convex hull of visited places	1
	Average trip speed	1
	Average trip distance	1
	Number of distinct places visited per day for each day type (weekday, Wednesday, weekend)	3
Mobility rhythms	Spatial entropy of visited locations	1
	Percentage of trips involving the most visited place	1
	Number of trips per time slot (8 weekday / 2 weekend), by day type (weekday, Wednesday, weekend)	18
Geographic context	Socio-economic indicators of the most visited place	11
	Local facilities/services near other visited places	24

While most indicators are directly computed from trajectories, geographic context variables required country-specific sources. French indicators rely on *French National Institute of Statistics and Economic Studies* census data, whereas Swiss indicators use StatPop and StatBL registries (Federal Statistical Office). Both provide comparable socio-economic data at a 200m grid resolution, from which 11 shared indicators (population, household types, housing characteristics) were extracted. Finally, local service accessibility was standardized across both domains using OpenStreetMap with identical tags and buffer logic.

Modeling pipeline

To predict socio-demographic attributes from aggregated mobility features, we employ the Extreme Gradient Boosting algorithm (XGBoost) following the methodology established in our previous work (Casassa et al., 2026). XGBoost was selected for its robustness, its strong predictive performance, and relatively low computational cost for hyperparameter tuning (Chen & Guestrin, 2016).
The inference task is framed as a supervised classification problem for each of the four target attributes : age (multi-class) and gender, diploma level, presence of children (all binary).
Formally, for each individual, weekly mobility patterns are encoded into a fixed-size feature vector $x \in \mathbb{R}^d$. The objective is to estimate a mapping function $f : \mathbb{R}^d \to \Omega$ for each attribute, where Ω represents the set of possible classes. The inference pipeline is illustrated in figure 2.

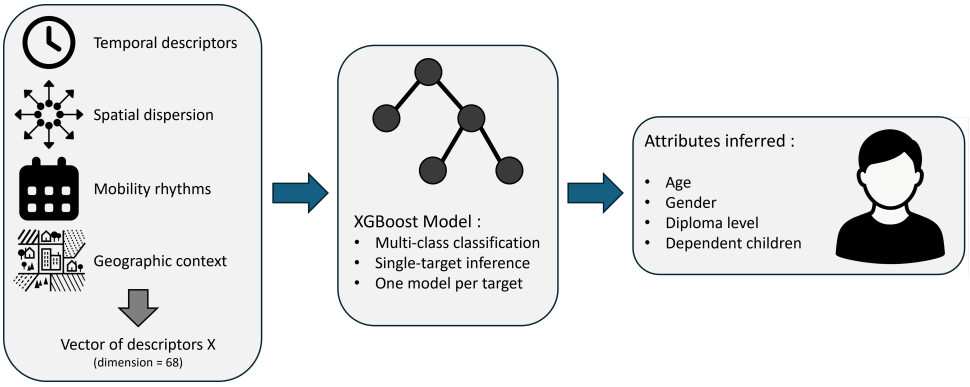


Figure 2: Inference pipeline overview. Trajectories are encoded into aggregated fixed-size vectors of descriptors. These features serve as inputs for independent XGBoost classifiers, where a specific model is trained for each target attribute using multi-class inference.

A central challenge of this study is to evaluate the model's portability across a source domain $\mathcal{D}_S$ (France) and a target domain $\mathcal{D}_T$ (Switzerland). We specifically investigated whether a function f_S optimized with the French dataset can accurately approximate the target function f_T in the Swiss domain. To this end, two experimental configurations were implemented:

- Intra-domain evaluation: Models were trained and evaluated within each country (70/30 split). Hyperparameters were optimized via 5-fold cross-validation, and class imbalance was handled using inverse-frequency weighting.
- Cross-domain transfer: The model trained on the French dataset is directly evaluated on the Swiss TimeUse+ dataset without further local adaptation.

Model performance is evaluated using the weighted F1-score, to account for class imbalances. For further details on the evaluation metrics, see (Casassa et al., 2026).

Quantification algorithms

While XGBoost classifiers provide individual-level predictions, directly aggregating these results yields biased population estimates due to dataset shifts in cross-domain transfer. To fulfill our second objective, which is obtaining socio-demographic distributions of the target domain, we transition from classification to quantification. The goal is to estimate the class prevalence $P(Y)$ in the Swiss target using French-trained classifiers. Thus, we evaluate three families of quantifiers that reconstruct $P(Y)$ from the classifiers' outputs. For mathematical definitions, we refer to Esuli

et al. (2023), while our implementation relies on the *QuantificationLib* library (Castaño et al., 2024).

- Aggregative Methods (CC & PCC): Classify & Count (CC) aggregate individual model outputs to estimate population-level proportions. Probabilistic Classify & Count (PCC) averages the posterior probabilities to better capture model uncertainty.
- Corrective Methods (AC & PAC): Adjusted Classify & Count (AC) and its probabilistic version (PAC) use a confusion matrix $\mathbf{C}$ (estimated from validation data) to invert observed bias, solving the linear system $\hat{P}_{corrected} = \mathbf{C}^{-1}\hat{P}_{aggregative}$ where $\hat{P}_{aggregative}$ represents the raw estimates from CC or PCC methods.
- Iterative Methods (EM): The Expectation-Maximization quantifier iteratively adjusts the model's posterior probabilities. By maximizing the likelihood of the observed scores in the target domain, it re-weights the initial predictions until the aggregate class prevalence converges.

To assess the accuracy of the quantifiers, the Mean Absolute Error (MAE) is evaluated (Esuli et al., 2023). It is defined as:

$$MAE(P, \hat{P}) = \frac{1}{|\Omega|} \sum_{y \in \Omega} |P(y) - \hat{P}(y)| \tag{1}$$

where P is the ground-truth distribution while $\hat{P}$ is the estimate distribution. The MAE yields 0 for a perfect estimate and reaches its maximum of $\frac{2}{|\Omega|}$ in case of total mismatch.

3 Results and discussion

This section evaluates the model's performance through a three-step analysis: (i) establishing intra-domain benchmarks and measuring direct transfer decay, (ii) characterizing the underlying dataset shifts and (iii) evaluating quantification algorithms to mitigate these biases and recover accurate class prevalence in the target domain.

Baseline and direct model transfer

Before assessing portability, we establish a performance baseline by training and testing XGBoost independently on the French (EMG) and Swiss (TimeUse+) datasets. This step serves to quantify the "upper bound" of predictability for each territory.

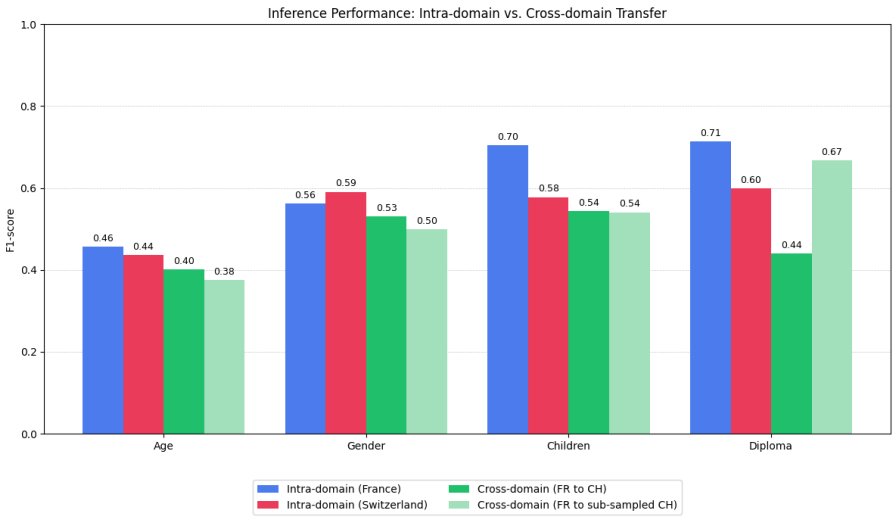


Figure 3: Model performance (F1-score) across attributes for four scenarios: intra-domain France, intra-domain Switzerland, direct cross-domain transfer (FR to CH), and cross-domain transfer on a sub-sampled Swiss population (aligned with French priors).

The intra-domain results demonstrate the effectiveness of mobility descriptors in both contexts. Figure 3 summarizes the F1-scores for each attribute, while the normalized confusion matrices in Figure 4 detail the specific recall and classification patterns within each country.
Intra-domain performance results reveal varied predictive capacities. For age, performance is consistent ($F1 \approx 0.45$), suggesting similar life-stage mobility signatures. Gender inference is robust in Switzerland ($F1 = 0.59$), while the French model struggles slightly with identifying men ($F1 = 0.56$). A notable disparity exists for dependent children, where French patterns appear more distinct ($F1 = 0.70$) than Swiss ones ($F1 = 0.58$). Finally, for diploma, the high French F1-score (0.71) is driven by a 75% majority class, whereas the Swiss model (0.60) shows better class balance.
Beyond these statistical considerations, the observed performance gaps between the two countries can also be tied to the nature of the descriptors themselves. The choice of descriptors, particularly the temporal segmentation of the day, was initially optimized based on the French mobility rhythms and cultural habits.
Despite this, the descriptors seem resilient enough to capture demographic signatures in both territories. These results align with the state-of-the-art results and benchmarks detailed in our previous work Casassa et al. (2026) (see section 5.3 for a comprehensive literature comparison).

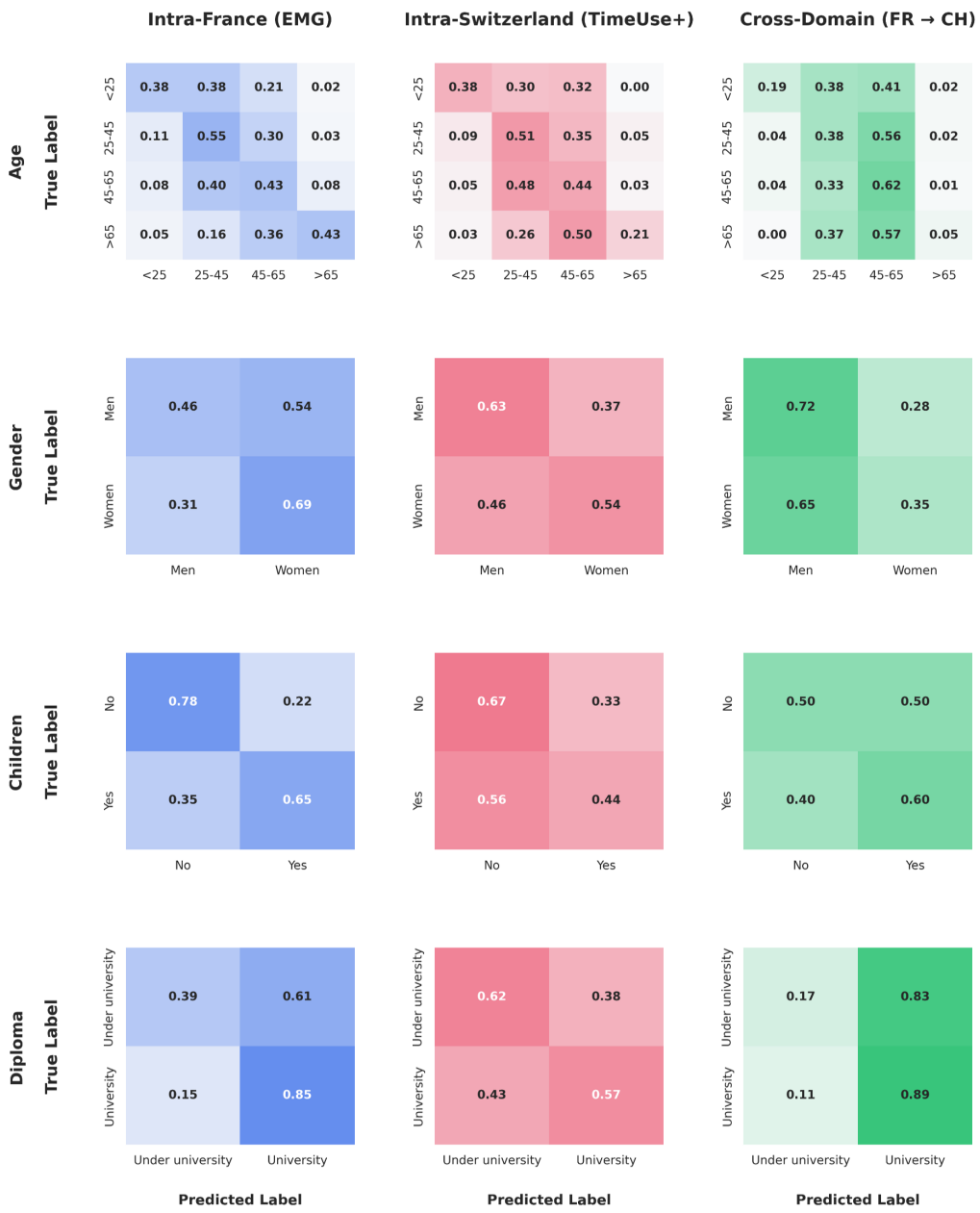


Figure 4: Normalized confusion matrices (row-wise) comparing local inference in France and Switzerland with cross-domain transfer scenario.

However, while these descriptors prove relevant within each country, their portability across domains presents a far more complex challenge. Direct model transfer reveals a sharp performance decay, as illustrated by the F1-score (green bars) in figure 3 and the green normalized confusion matrices in figure 4.
Performance decline varies by attribute. For age, gender, and education, the model exhibits clear overconfidence, systematically biasing predictions toward a single dominant category. Notably, it misclassifies 83% of Swiss "below university" profiles. Conversely, parenthood inference shows a near-random 50/50 split. This maximal uncertainty suggests that mobility patterns associated with having children are territory-specific and fail to generalize from France to Switzerland.
In conclusion, the results reveal a substantial degradation in performance under cross-domain transfer. Although models trained and evaluated within the same country achieve comparable performance, direct transfer leads to a marked collapse in discriminative ability, often failing to surpass random or majority-class baselines. Remaining anchored to the decision boundaries and class priors learned in France, the model effectively reflects the structure of its training domain rather than adapting to the target context.

Characterization of dataset shifts

To explain these predictive biases, our objective is to characterize the impact of the three distinct dataset shifts mentioned earlier : prior probability shift, covariate shift, and concept shift. Prior probability shift, occurs when the distribution of the target classes $P(Y)$ differs between the source and target domains. The availability of ground truth labels for the Swiss dataset allows for a straightforward validation of this shift, as illustrated in Figure 5.
The comparison reveals structural differences in population samples, most notably for the diploma variable: higher education accounts for more than 75% of the French sample but only 47% of the Swiss one. While age distributions are consistent for middle categories, the French sample is skewed toward the youngest group. Regarding gender, the French dataset slightly favors women, whereas the Swiss sample is more balanced toward men. Finally, the Swiss sample shows a marginally lower proportion of individuals with children compared to the French baseline.

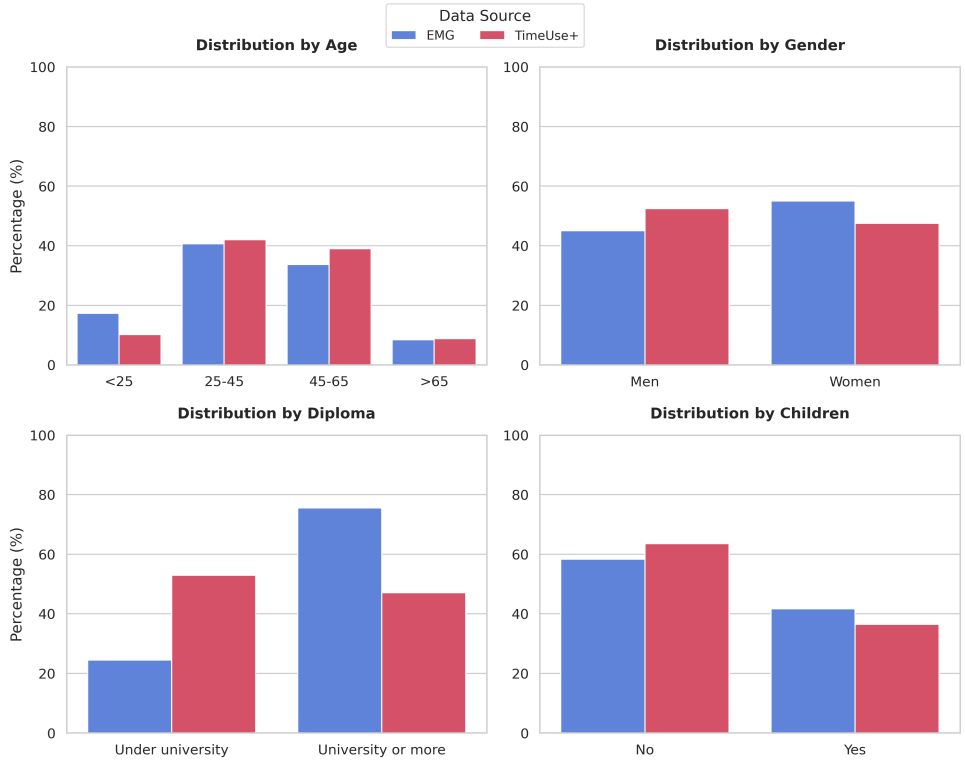


Figure 5: Socio-demographic distributions of the French and Swiss datasets.

The covariate shift occurs when the marginal distribution of the input features $P(X)$ changes between domains. To quantify this divergence between the French and Swiss datasets, we employed two complementary approaches:

- Projecting the 68 descriptors onto a shared 2D space reveals a relatively clear separation between the two populations, although they partially overlap (figure 6). The horizontal axis (PC1, 21.0% variance) primarily captures environmental and infrastructure differences, such as accessibility to public transport and store density. The vertical axis (PC2, 8.2%) discriminates the duration of the trip and the weekly activity count.

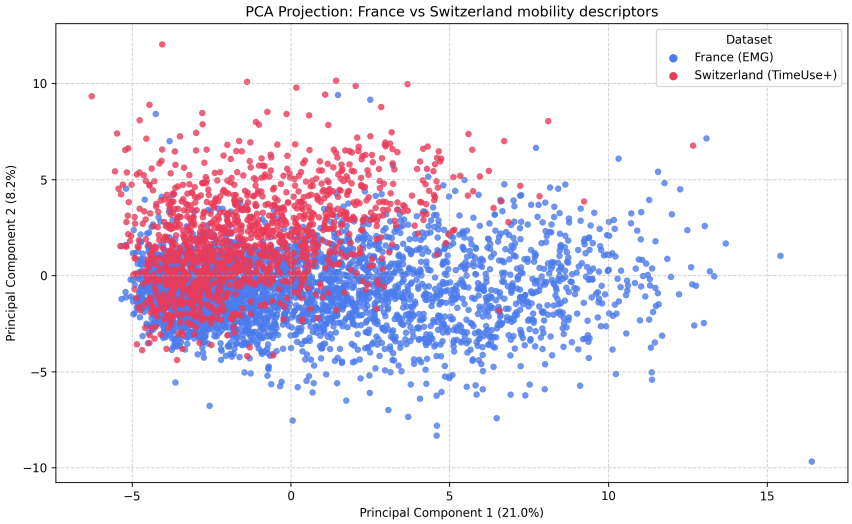


Figure 6: Two-dimensional PCA projection of mobility descriptors: France vs. Switzerland.

- Domain classifier: To quantify this shift, a Random Forest was trained to predict the country of origin of the feature vector. The model achieved a near-perfect accuracy of 0.96, proving that French and Swiss mobility signatures are statistically distinct. Feature importance identifies average speed, temporal habits (notably morning trip frequency and earliest departure times), and public transport use as the most prominent markers of difference.

By examining in detail the descriptors ranked as the most discriminant by Random Forest, covariate shift seems particularly pronounced for several descriptors (figure 7). For instance, Swiss mobility is characterized by higher average speeds and an earlier daily start, with a sharp peak at 5 a.m. These divergences are amplified by environmental contrasts in population density and service accessibility. However, some structural indicators, such as the number of weekly POIs (Points Of Interest) or the proportion of trips to primary locations, remain stable across domains, suggesting certain universal behavioral mechanisms.
The critical challenge remains concept shift, where the feature-label relationship diverges. In order to partially isolate this effect, we decompose the joint distribution as $P(X, Y) = P(X|Y)P(Y)$. We neutralized prior shift by sub-sampling the Swiss dataset to match French class proportions $P(Y)$. Despite this alignment, a persistent performance gap (figure 3, light green) confirms that the drop arises from changes in $P(X|Y)$, highlighting the joint impact of covariate and concept shifts. In this sub-sampled scenario, F1-scores are slightly lower than in the direct transfer, likely due to the reduced sample size and loss of diversity during re-sampling. An exception is the diploma attribute ($F1 = 0.67$), which is artificially inflated by a strong classification bias.
To summarize, this decline is not merely due to differing class proportions. PCA and domain classification confirm a massive covariate shift, driven by distinct infrastructure and cultural mobility rhythms. A significant concept shift also appears to be present, although it remains difficult to disentangle from the effects of covariate shift.

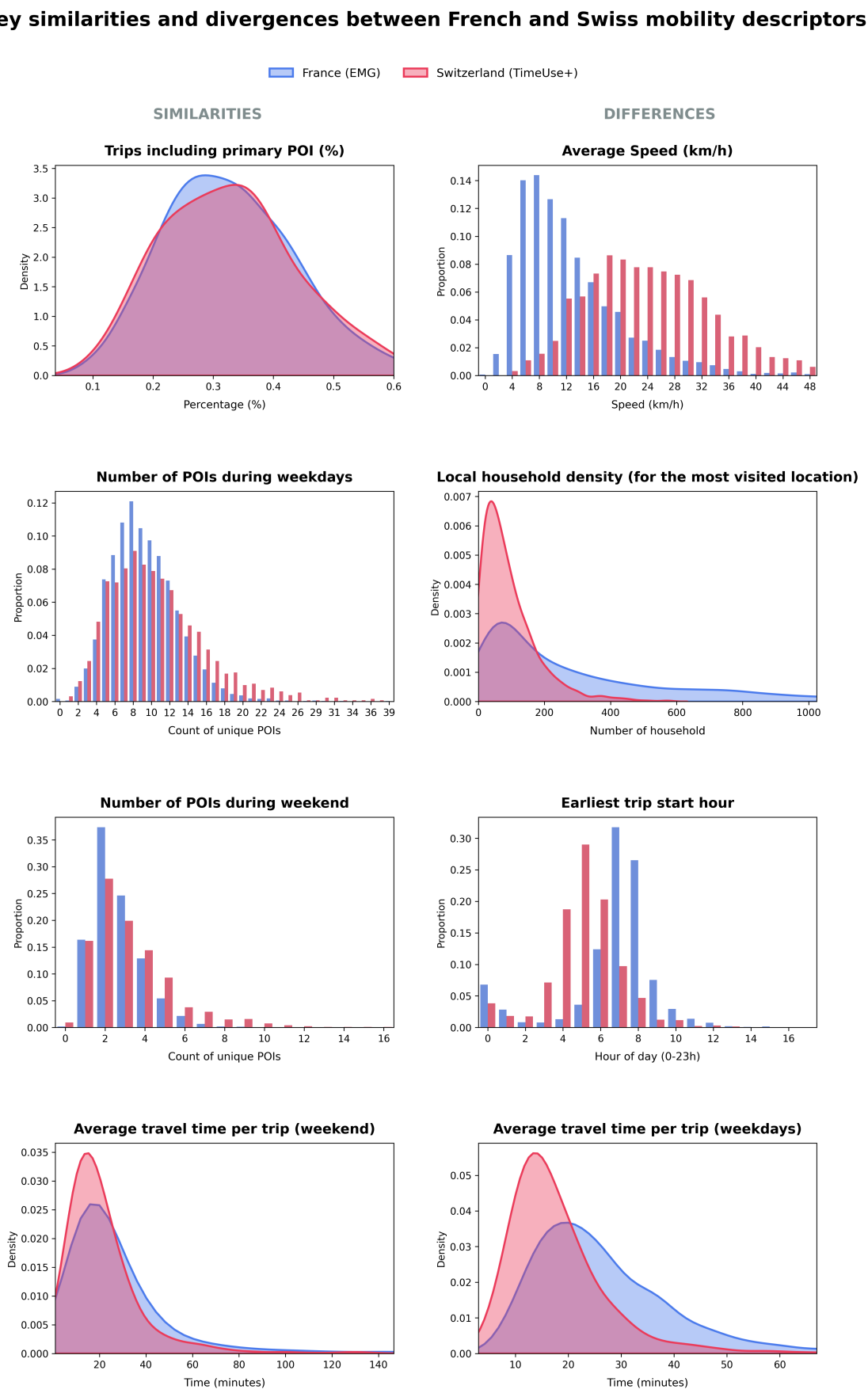


Figure 7: Comparative density and proportion analysis of key mobility predictors in France and Switzerland.

Quantification

Although the preceding analysis demonstrates that all three types of dataset shifts coexist, this section investigates whether quantification algorithms can adjust the bias of classifiers outputs to recover global distribution estimates. As these methods primarily assume a pure prior shift, the divergence in conditional distributions $P(X|Y)$ between domains suggests that, as could be expected, they may struggle in our specific setting. Tables 2 to 5 summarize the performance of the framework described in the methodology.

The primary observation is the near-total collapse of advanced corrective methods such as AC, PAC, and EM. These algorithms produce extreme estimates, converging towards 0 or 1 for specific classes. While the low condition numbers of the confusion matrices ($\kappa < 55$) for corrective methods (AC and PAC) confirm that this failure is not due to non-invertibility, it suggests that the dataset shifts exceed the methods' corrective capacities. An exception is observed for the PAC method regarding the diploma attribute, which yields a very high-quality result (0.033).

The Expectation-Maximization (EM) quantifier also fails to converge toward meaningful values. This collapse may be attributed to the initialization of the algorithm. Since EM iteratively updates the prevalences starting from the French training priors, the significant gap between the French and Swiss datasets likely causes the likelihood function to fall into a local minimum.

Interestingly, the Probabilistic Classify and Count (PCC) method consistently yields the lowest MAE error for age (0.065), gender (0.021), children (0.157) and the second lower score for diploma (0.138). PCC manages to recover distributions that remain reasonably close to the ground truth.

To conclude, quantification methods offer limited corrective power against such heterogeneous shifts. Interestingly PCC method emerged as the most resilient tool. This suggests that despite biased discrete classifications, posterior probabilities retain a meaningful global signal that can be effectively leveraged to estimate population-level distributions.

Table 2: Quantification results for Age.

Method	**<25**	**25-45**	**45-65**	**>65**	**MAE**
Truth	**0.102**	**0.420**	**0.390**	**0.088**	**0.000**
CC	0.051	0.360	0.568	0.021	0.089
PCC	**0.164**	**0.325**	**0.354**	**0.157**	**0.066**
AC	0.105	0.249	0.281	0.366	0.140
PAC	0.000	0.000	1.000	0.000	0.305
EM	0.000	0.000	0.000	1.000	0.456

Table 3: Quantification results for Gender.

Method	**Men**	**Women**	**MAE**
Truth	**0.525**	**0.475**	**0.000**
CC	0.683	0.317	0.158
PCC	**0.546**	**0.454**	**0.021**
AC	1.000	0.000	0.475
PAC	1.000	0.000	0.475
EM	1.000	0.000	0.475

Table 4: Quantification results for Children.

Method	**No**	**Yes**	**MAE**
Truth	**0.635**	**0.365**	**0.000**
CC	0.461	0.539	0.174
PCC	**0.478**	**0.522**	**0.157**
AC	0.167	0.833	0.468
PAC	0.202	0.798	0.433
EM	0.066	0.934	0.569

Table 5: Quantification results for Diploma.

Method	Under university	University	MAE
Truth	**0.529**	**0.471**	**0.000**
CC	0.145	0.855	0.385
PCC	0.392	0.608	0.138
AC	0.036	0.964	0.494
PAC	**0.497**	**0.503**	**0.033**
EM	1.000	0.000	0.470

4 Conclusions

This study addresses our research questions by highlighting the significant performance collapse in individual-level inference when our model is directly transferred between the French and Swiss datasets, despite their geographical proximity. This decline stems from the joint impact of three distinct dataset shifts: prior, covariate, and concept shifts. Quantification methods offer an alternative for recovering global population distributions. However, these methods remain limited as they operate under the restrictive assumption of a pure prior shift.

To move beyond these limitations and the inherent bias of descriptors tailored to French mobility patterns, we plan to explore deep learning architectures such as LSTMs and GNNs. Building on our previous work (Casassa et al., 2026), these models extract representations directly from raw trajectories, without manual feature engineering. Leveraging these learned representations, we also intend to investigate the reverse transfer scenario (Switzerland to France).

Moreover, since covariate and concept shifts also significantly distort the feature-label relationship across domains, future research will focus on more robust domain generalization frameworks. In particular, we aim to incorporate alignment mechanisms during training that can jointly account for variations in both $P(Y)$ and $P(X|Y)$ (Yang et al., 2022; Yeom et al., 2026).

While individual-level adjustment strategies represent a valuable direction for future work to enhance user profile enrichment, this study focuses on population-level reconstruction. This global perspective enables accurate reconstruction from anonymous datasets, providing a strategic lever to assess their statistical representativeness. Ultimately, this approach can help unlock the potential of large-scale trajectory data as a reliable and cost-effective complement to traditional surveys for transport planning.

Acknowledgements

The authors would like to thank the *Institut Paris Région* and *ETH Zurich* for granting access to the datasets that made this research possible.

References

Almaatouq, A., Prieto-Castrillo, F., & Pentland, A. (2016). Mobile communication signatures of unemployment. In *Social informatics* (pp. 407–418).

Bakhtiari, A., Mirzahossein, H., Kalantari, N., & Jin, X. (2023). Inferring socioeconomic characteristics from travel patterns. *Journal of Regional and City Planning*, *34*, 122–136.

Casassa, E., Côme, E., Oger, R., & Oukhellou, L. (2026, June). Inferring socio-demographic attributes from weekly gps trajectories: a case study in the paris region. In *Proceedings of the 17th world conference on transport research (2026).* Toulouse, France. (Accepted)

Castaño, A., Alonso, J., González, P., Pérez, P., & del Coz, J. J. (2024). Quantificationlib: A python library for quantification and prevalence estimation. *SoftwareX*, *26*, 101728.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Esuli, A., Fabris, A., Moreo Fernández, A., & Sebastiani, F. (2023). *Learning to quantify.*

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

Meister, A., Schoenduwe, R., Tirban, P., Winkler, C., Schmid, B., & Axhausen, K. (2020). *Timeuse+: A multi-layered travel and time-use diary format* (Arbeitsbericht Verkehrs und Raumplanung No. 1543). ETH Zurich.

Moreno-Torres, J. G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N. V., & Herrera, F. (2012). A unifying view on dataset shift in classification. *Pattern Recognition*, *45*(1), 521–530.

Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (2009). *Dataset shift in machine learning.* Cambridge, MA: The MIT Press.

Schumacher, T., Strohmaier, M., & Lemmerich, F. (2025). *A comparative evaluation of quantification methods.*

Stopher, P., & Greaves, S. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, *41*, 367–381.

Winkler, C., Meister, A., & Axhausen, K. (2024). The timeuse+ data set: 4 weeks of time use and expenditure data based on gps tracks. *Transportation*, *53*, 1–27.

Wu, L., Yang, L., Huang, Z., Wang, Y., Chai, Y., Peng, X., & Liu, Y. (2019). Inferring demographics from human trajectories and geographical context. *Computers, Environment and Urban Systems*, *77*, 101368.

Yang, Y., Wang, H., & Katabi, D. (2022). *On multi-domain long-tailed recognition, imbalanced domain generalization and beyond.*

Yeom, J., Chae, K., Lim, H., Oh, Y., Yang, D., & Kim, T. (2026). *Robust domain generalization under divergent marginal and conditional distributions.*