

# Hybrid Simulator for Capturing Dynamics of Synthetic Population

Marija Kukic   Salim Benchelabi   Michel Bierlaire

**STRC**

**23rd Swiss Transport Research Conference**  
Monte Verità / Ascona, May 10 – 12, 2023

May 10, 2023



**EPFL**

# Outline

- 1 Motivation
- 2 Literature review
- 3 Methodology
- 4 Results: Case study of Switzerland
- 5 Conclusion and Future Work

# Synthetic population: What? Why?

## Real Data

- High cost of data collection.
- Lack of representativity.
- Data privacy constraints.

## Synthetic Data

- Open source.
- Bias correction.
- Privacy preservation.

# Synthetic population: What? Why?

## Real Data

- High cost of data collection.
- Lack of representativity.
- Data privacy constraints.

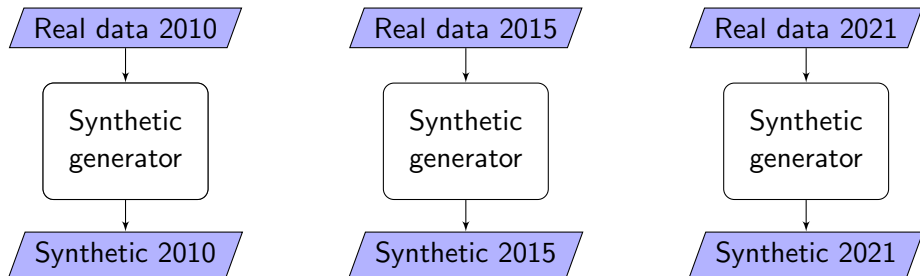
## Synthetic Data

- Open source.
- Bias correction.
- Privacy preservation.

## Synthetic Population in Transportation?

**Generation algorithms:** Statistical reconstruction, Combinatorial Optimization and Statistical learning

# Synthetic population: Snapshot of the data



## Problems: Independent re-generation?

- Complicated and costly re-generation.
- No use of the past data and all available data sources.
- Outdated synthetic population.

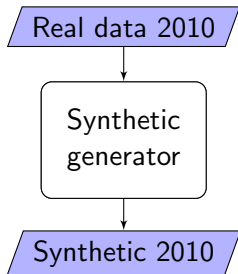
Population evolves over time = **Capturing Dynamics**

How to capture dynamics? => **Projection of generated sample**

# Outline

- 1 Motivation
- 2 Literature review**
- 3 Methodology
- 4 Results: Case study of Switzerland
- 5 Conclusion and Future Work

# Literature review: Generation and Projection



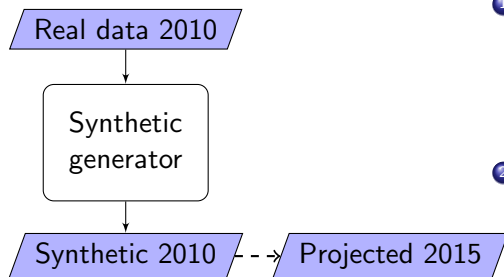
Step 1: Generation

- 1 **Statistical Reconstruction**  
[Fatmi and Habib, 2017,  
Prédhumeau and Manley, 2023]
- 2 **Combinatorial Optimization**  
[Namazi-Rad et al., 2014]

**Statistical learning?**



# Literature review: Generation and Projection



## ① Dynamic Projection

Simulate life events

[Namazi-Rad et al., 2014,  
Fatmi and Habib, 2017]

## ② Re-sampling

Adjust marginals

[Prédhumeau and Manley, 2023]

Step 1: Generation    Step 2: Projection

# Literature review: Gaps

## Problems of projection

- Arbitrarily chosen choice of the generators.
- Limited number of considered attributes.
- Lack of validation.

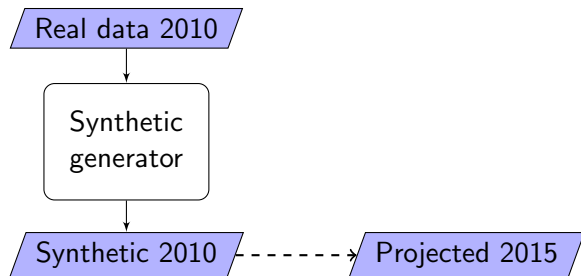
## Dynamic projection

- Propagation of the generation bias and errors.
- Increase of the error over time.
- Not robust to the unusual events.

## Re-sampling

- Lack of heterogeneity.

# Contribution: Hybrid Simulator for Capturing Dynamics

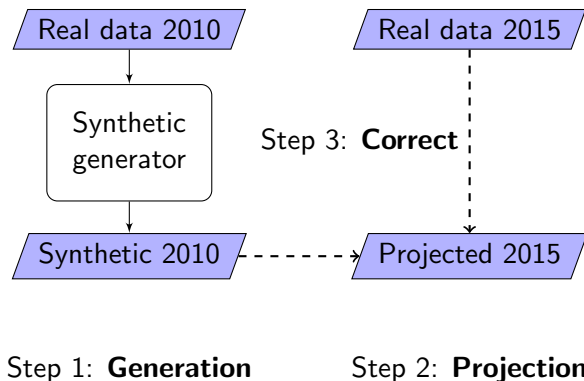


Step 1: Generation

Step 2: Projection

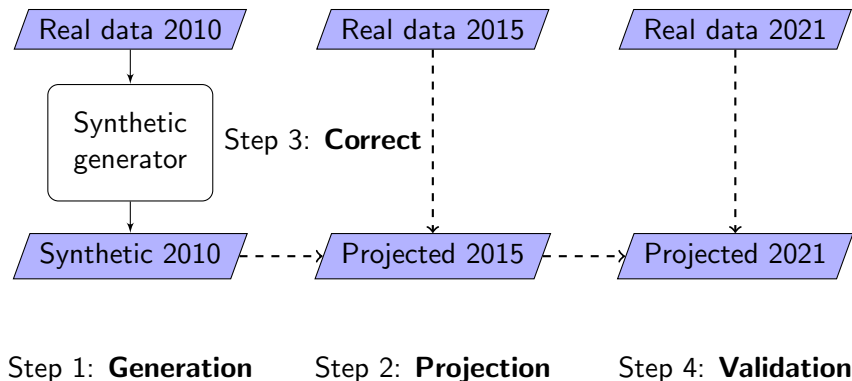
**Model-based approach**

# Contribution: Hybrid Simulator for Capturing Dynamics



**Model-based and Data-driven approach**

# Contribution: Hybrid Simulator for Capturing Dynamics



**Model-based and Data-driven approach**

# Outline

- 1 Motivation
- 2 Literature review
- 3 Methodology**
- 4 Results: Case study of Switzerland
- 5 Conclusion and Future Work

# Hybrid Simulator for Capturing Dynamics

## Step 1: Generation

Markov Chain Monte Carlo Simulation [Kukic and Bierlaire, 2023]

Synthetic individuals  $X = (X_{\text{age}}, X_{\text{emp}}, X_{\text{gender}})$

Bootstrap and convergence monitoring

## Step 2: Dynamic projection

When disaggregated data are not available.

Simulate events: birth, death and migration.

Simulate impact on age, gender and employment.

# Hybrid Simulator for Capturing Dynamics

## Step 3: Re-sampling

When disaggregated data are available.

Compare age marginals with real data.

Add or delete individuals to achieve desired fit.

## Step 4: Validation

Compare marginal and sub-distributions with real data.

Statistics (e.g., SRMSE) and Visualization.

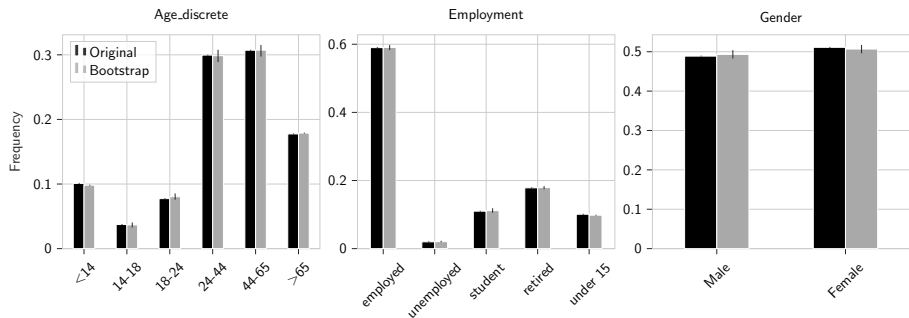


# Outline

- 1 Motivation
- 2 Literature review
- 3 Methodology
- 4 Results: Case study of Switzerland**
- 5 Conclusion and Future Work

# Generation and validation of synthetic sample - 2010

Reference data: weighted **MTMC 2010, 2015, 2021** [OFS]



**Figure:** The comparison of the marginal distributions between synthetic and real sample from 2010

# Dynamic Projection (2010 - 2014) and Re-sampling (2015)

Rates on **birth, death and migration** [OFS]

$\cdot 10^{-2}$

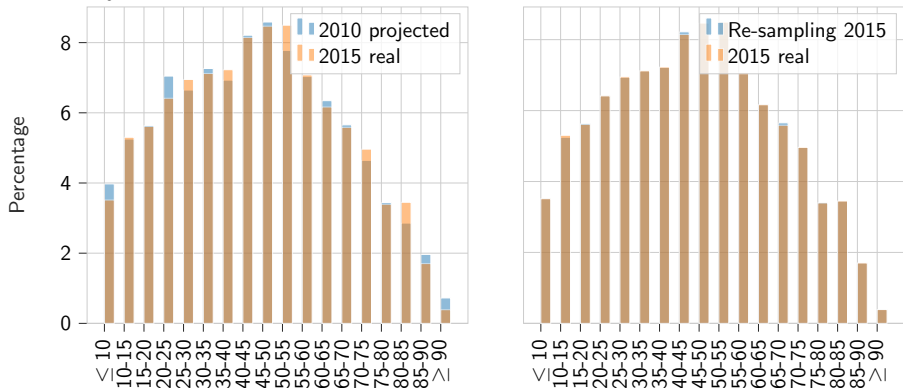


Figure: Comparison with real data 2015  
(Left - Projection results; Right - Re-sampling results)

# Comparison of projection and hybrid approach - 2021

	Age discrete	Employment	Gender	Average All attributes
Hybrid approach 2010 - 2021	<b>0.073</b>	<b>0.052</b>	<b>0.006</b>	<b>0.044</b>
Projection 2010 - 2021	0.082	0.071	0.006	0.053

**Table:** SRMSE of projected samples against real sample 2021

# Comparison of projection and hybrid approach - 2021

	Age discrete	Employment	Gender	Average All attributes
Projection <b>2015 to 2021</b>	<b>0.057</b>	<b>0.037</b>	<b>0.004</b>	<b>0.033</b>
Hybrid approach <b>2010 - 2021</b>	0.073	0.052	0.006	0.044
Projection <b>2010 - 2021</b>	0.082	0.071	0.006	0.053

**Table:** SRMSE of projected samples against real sample 2021

# Outline

- 1 Motivation
- 2 Literature review
- 3 Methodology
- 4 Results: Case study of Switzerland
- 5 Conclusion and Future Work**

# Conclusion and Future Work

## Summary

- Maintenance of synthetic samples without regenerating.
- Access to up-to-date data and making use of the past.
- Hybrid approach - trade-off between accuracy and efficiency.

## Independent generation VS. Hybrid approach

Number of attributes.

Availability of the real sample.





Re-sample other attributes than age.

# Thank you :) Questions?





# Bibliography

-  Fatmi, M. R. and Habib, M. A. (2017).  
Baseline synthesis and microsimulation of life-stage transitions within an agent-based integrated urban model.
-  Kukic, M. and Bierlaire, M. (2023).  
Divide-and-conquer one-step simulator for the generation of synthetic households.
-  Namazi-Rad, M.-R., Mokhtarian, P., and Perez, P. (2014).  
Generating a dynamic synthetic population – using an age-structured two-sex model for household dynamics.
-  Prédhumeau, M. and Manley, E. (2023).  
A synthetic population for agent based modelling in Canada.