# A new framework for assessing classification algorithms for mode choice prediction

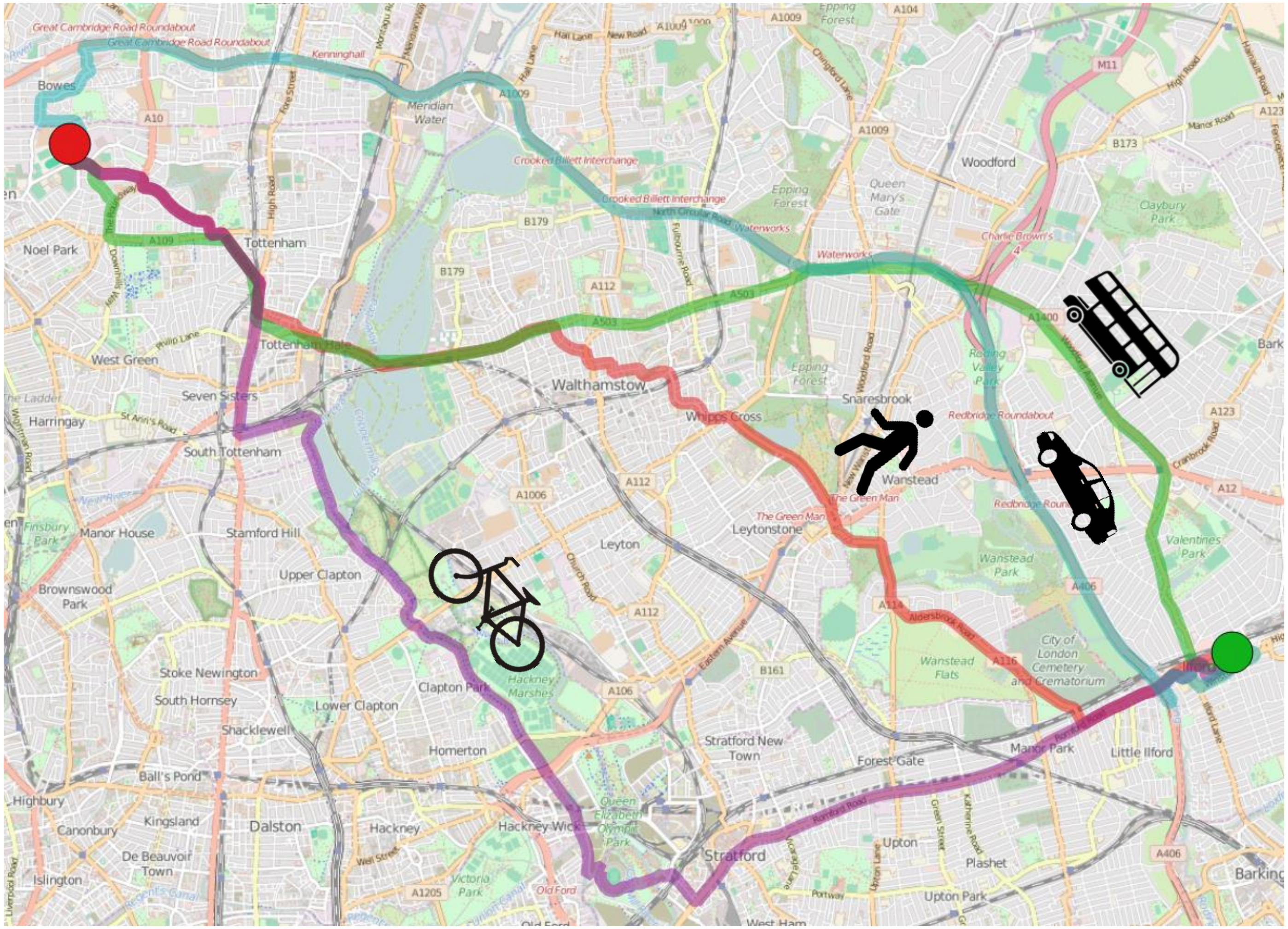
hEART 2018

Tim Hillel, Michel Bierlaire, Mohammed Elshafie, Ying Jin

University of Cambridge

École Polytechnique Fédérale de Lausanne EPFL

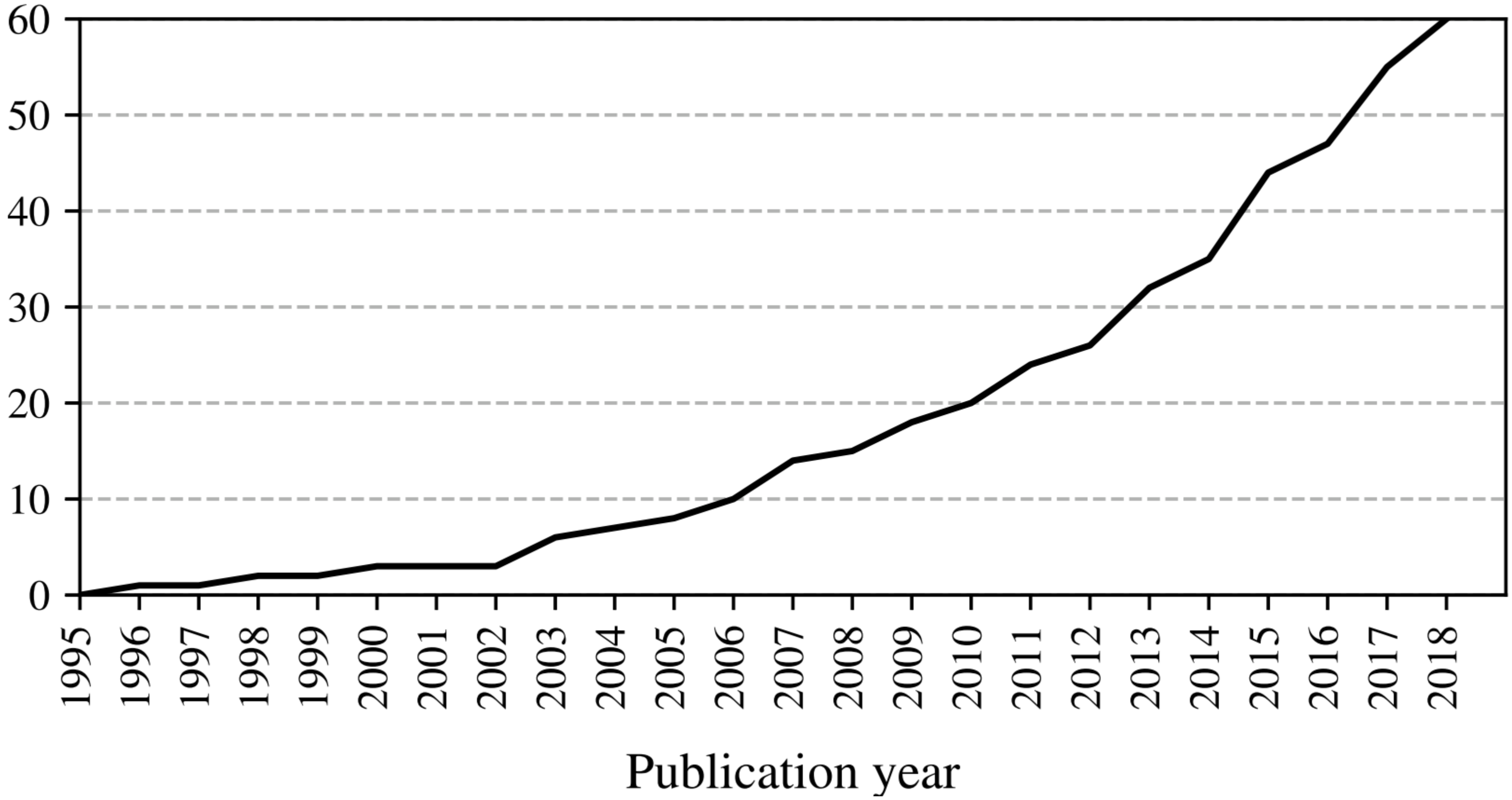
# Motivation




UNIVERSITY OF CAMBRIDGE

EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Machine Learning: Increasing research




UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Need for framework?

Systematic review:

*ML methodologies for mode-choice modelling*

# Need for framework?

Systematic review:

*ML methodologies for mode-choice modelling*

60 papers → 63 studies

# Motivation: Classification techniques

| No. | DCM | ANN | DT | EL | SVM | BL | RBML | HM | Msc |
|---|---|---|---|---|---|---|---|---|---|
| S1 | ✓ | | | | | | ✓ | | |
| S2 | ✓ | ✓ | | | | | | | |
| S3 | ✓ | | ✓ | ✓ | ✓ | ✓ | | | |

# Motivation: Technical issues

10 Technical limitations

# Motivation: Technical issues

10 Technical limitations

*Today only discuss 3*

# Overview: Technical issues

1. Treating choice prediction as deterministic

2. Using trip-wise sampling with hierarchical data

3. Not optimising model hyper-parameters

# Overview

1. For each issue
    a. Explanation
    b. Implementation in literature
    c. Proposed solution

2. Dataset and methodology

3. Results

# 1. Deterministic prediction

UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# 1. Deterministic prediction

84% (53 studies) use only deterministic metrics

UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# 1. Deterministic prediction

84% (53 studies) use only deterministic metrics

*100x less cyclists predicted using deterministic classification*

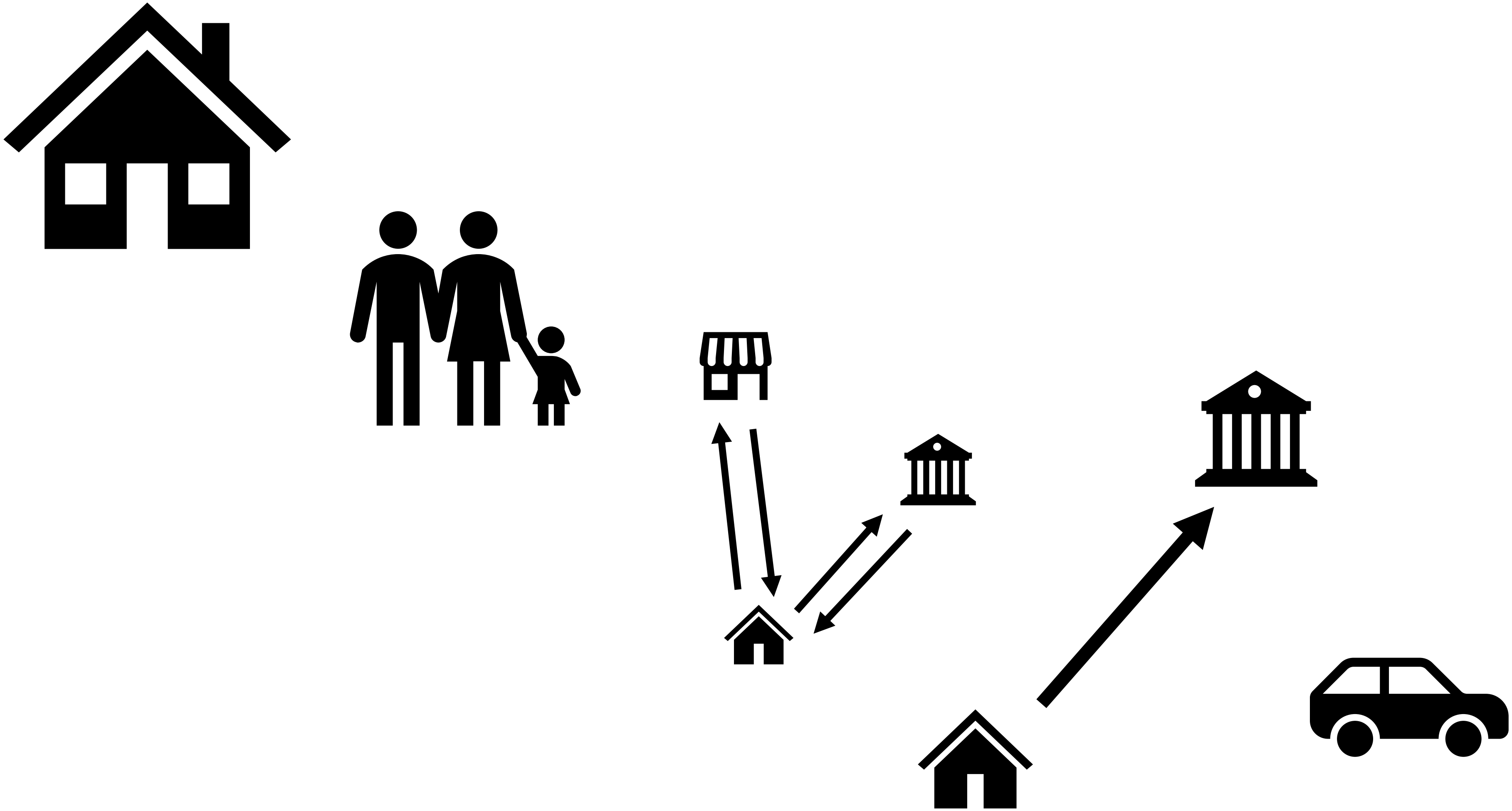
# 1. Deterministic prediction

84% (53 studies) use only deterministic metrics

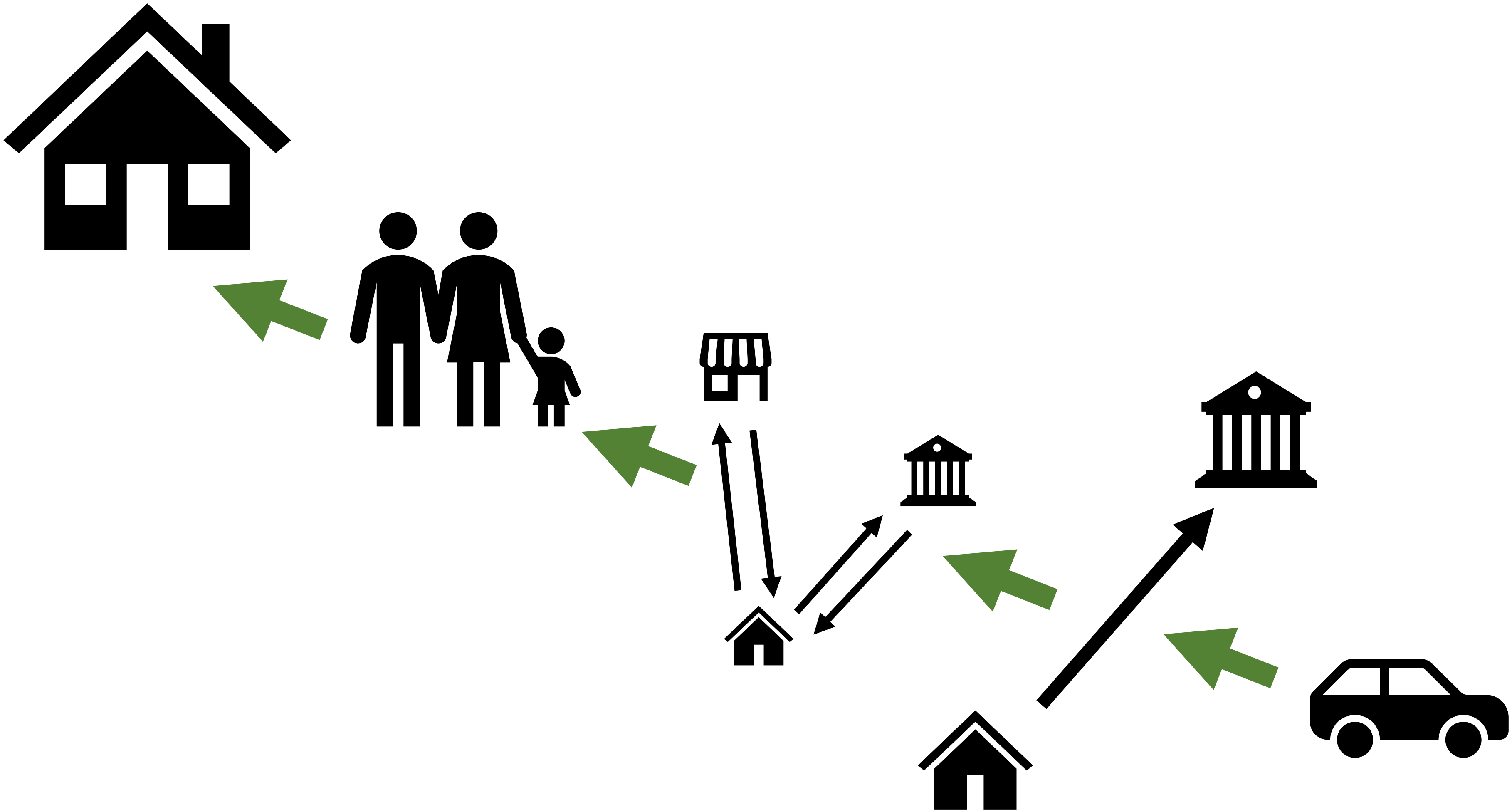
*Solution: use stochastic metrics (Log likelihood)*

# 1. Deterministic prediction

# 2. Hierarchical data



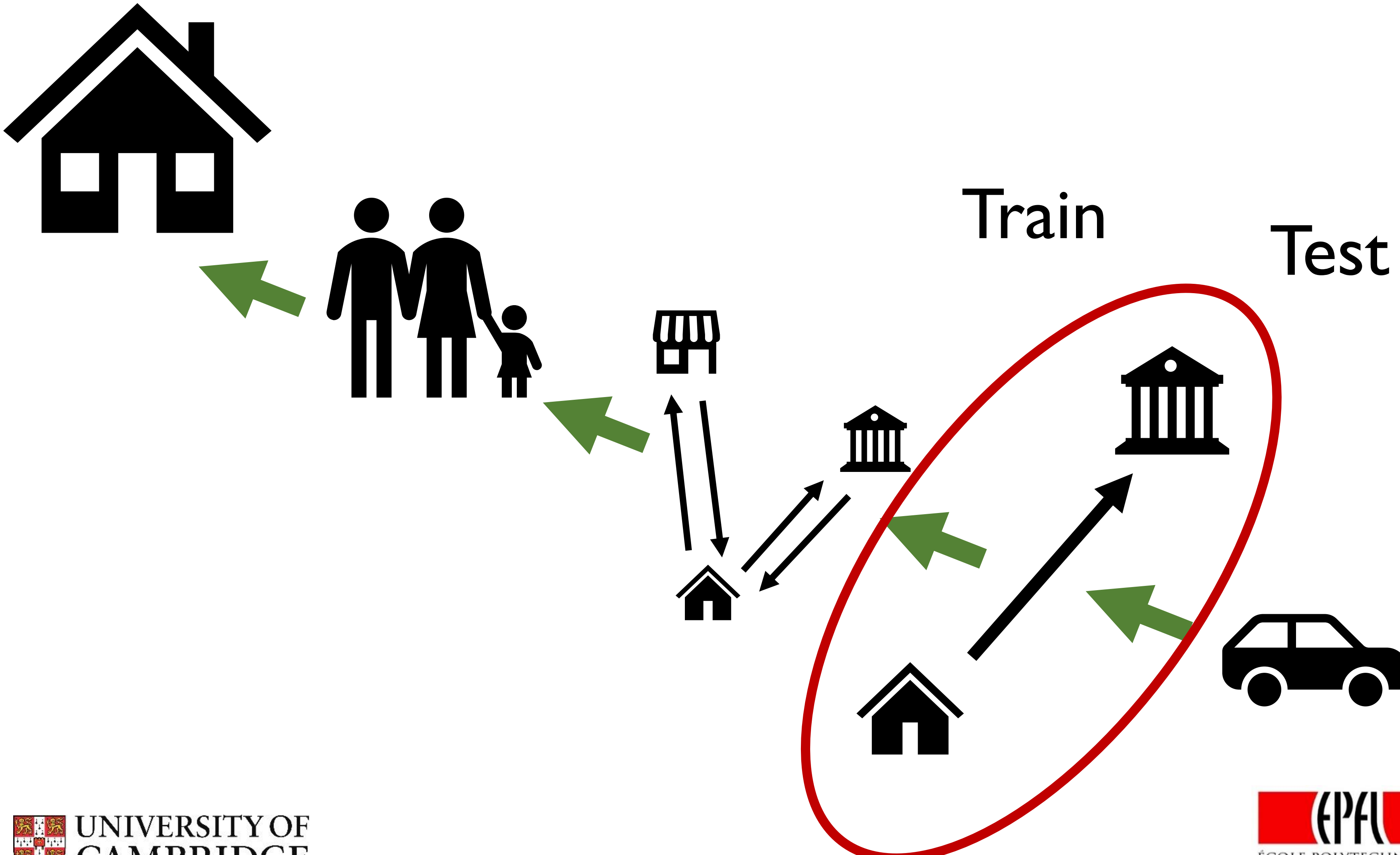
# 2. Hierarchical data



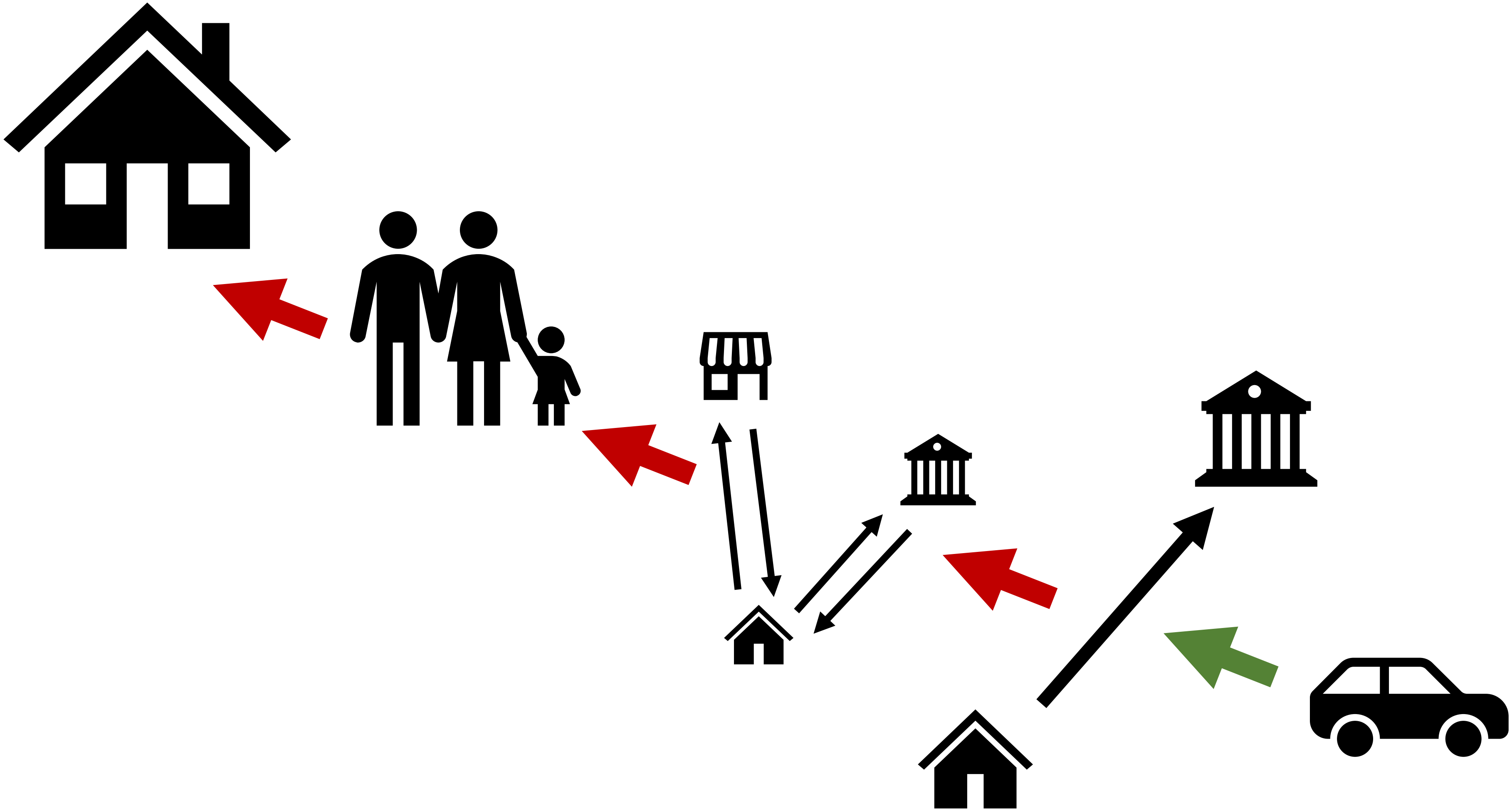
UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# 2. Hierarchical data



# 2. Hierarchical data



UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# 2. Hierarchical data

56% (35 studies) use hierarchical data

UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# 2. Hierarchical data

56% (35 studies) use hierarchical data

All use trip-wise sampling

# 2. Hierarchical data

56% (35 studies) use hierarchical data
All use trip-wise sampling

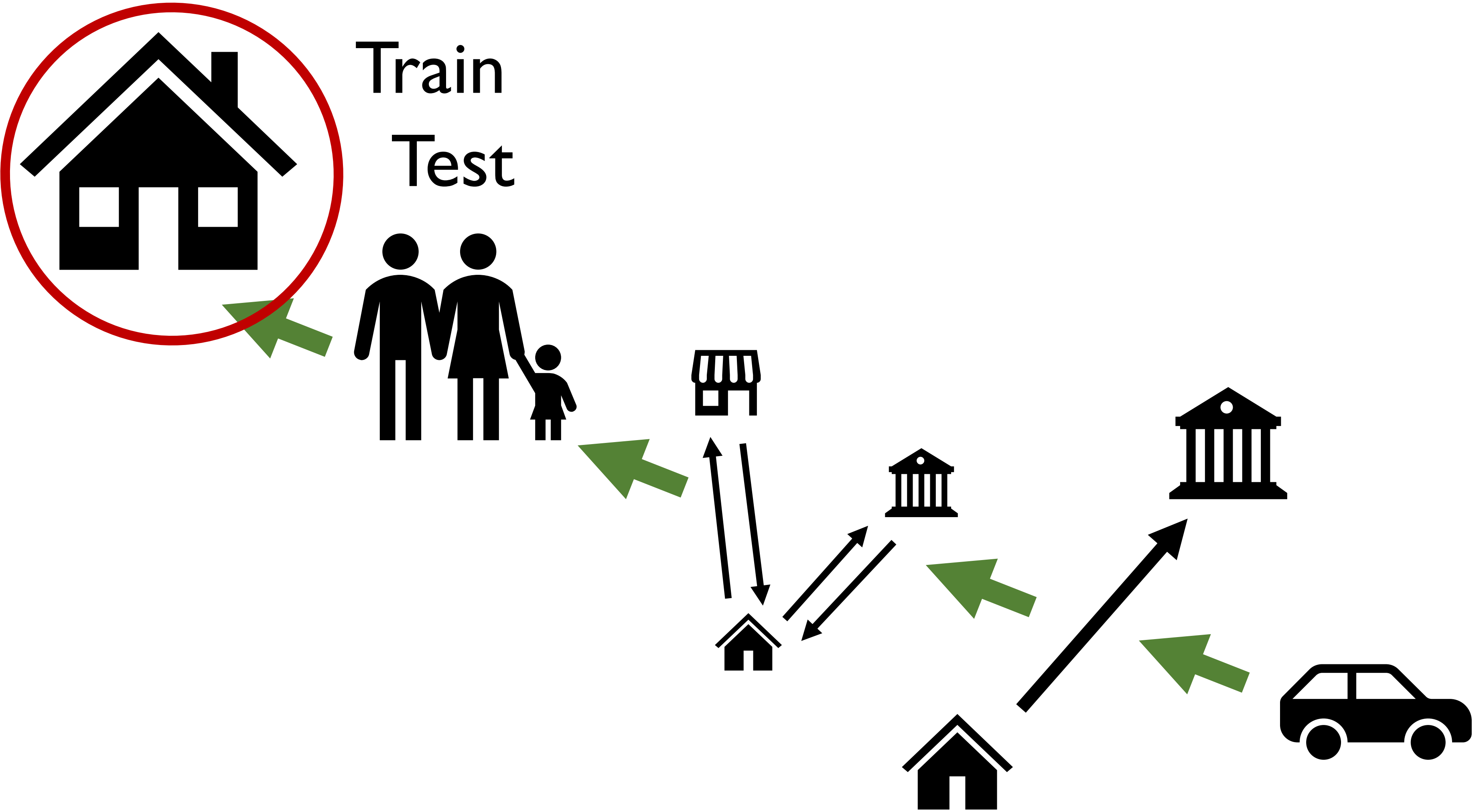
*1 day trip diary – 10 fold CV – 68% repeated trips with same mode*

# 2. Hierarchical data

56% (35 studies) use hierarchical data
All use trip-wise sampling

*Solution: household-wise sampling*

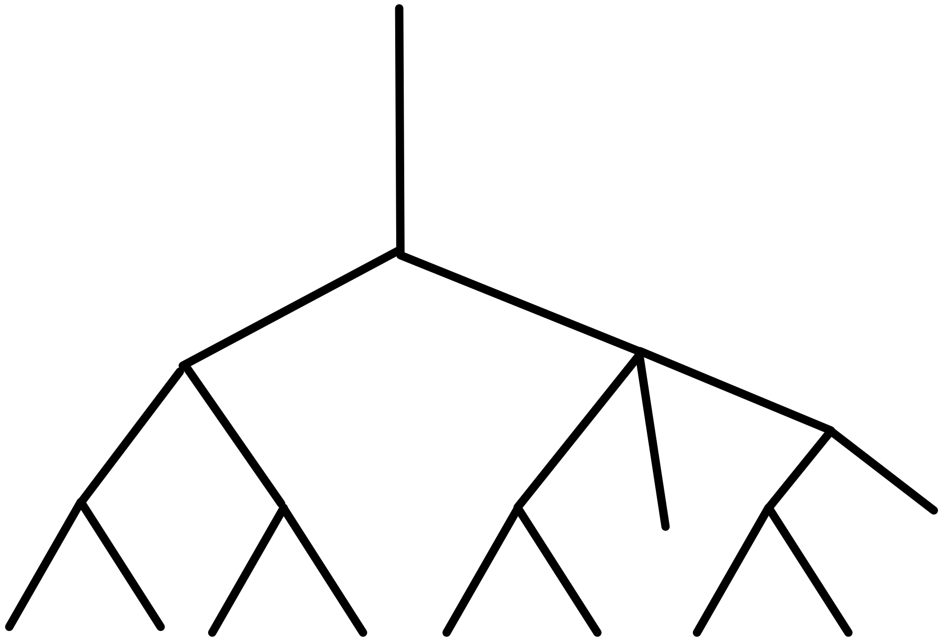
# 2. Hierarchical data




UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# 3. Optimising hyper-parameters



# 3. Optimising hyper-parameters

40% (20/49 studies) perform no hyperparameter optimisation

# 3. Optimising hyper-parameters

1/49 studies implement sequential optimisation

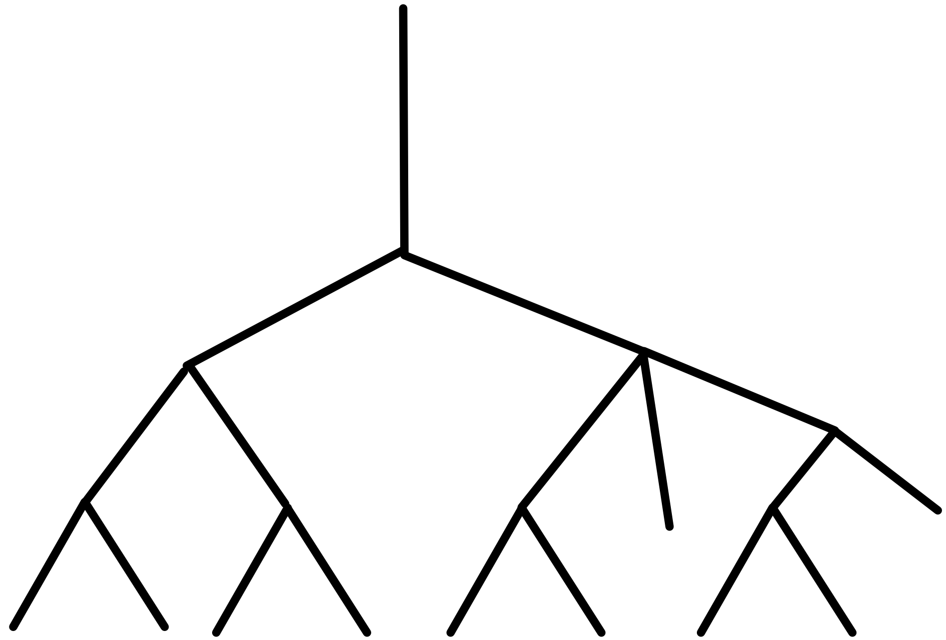
*Grid search performs worse than random search*

# 3. Optimising hyper-parameters

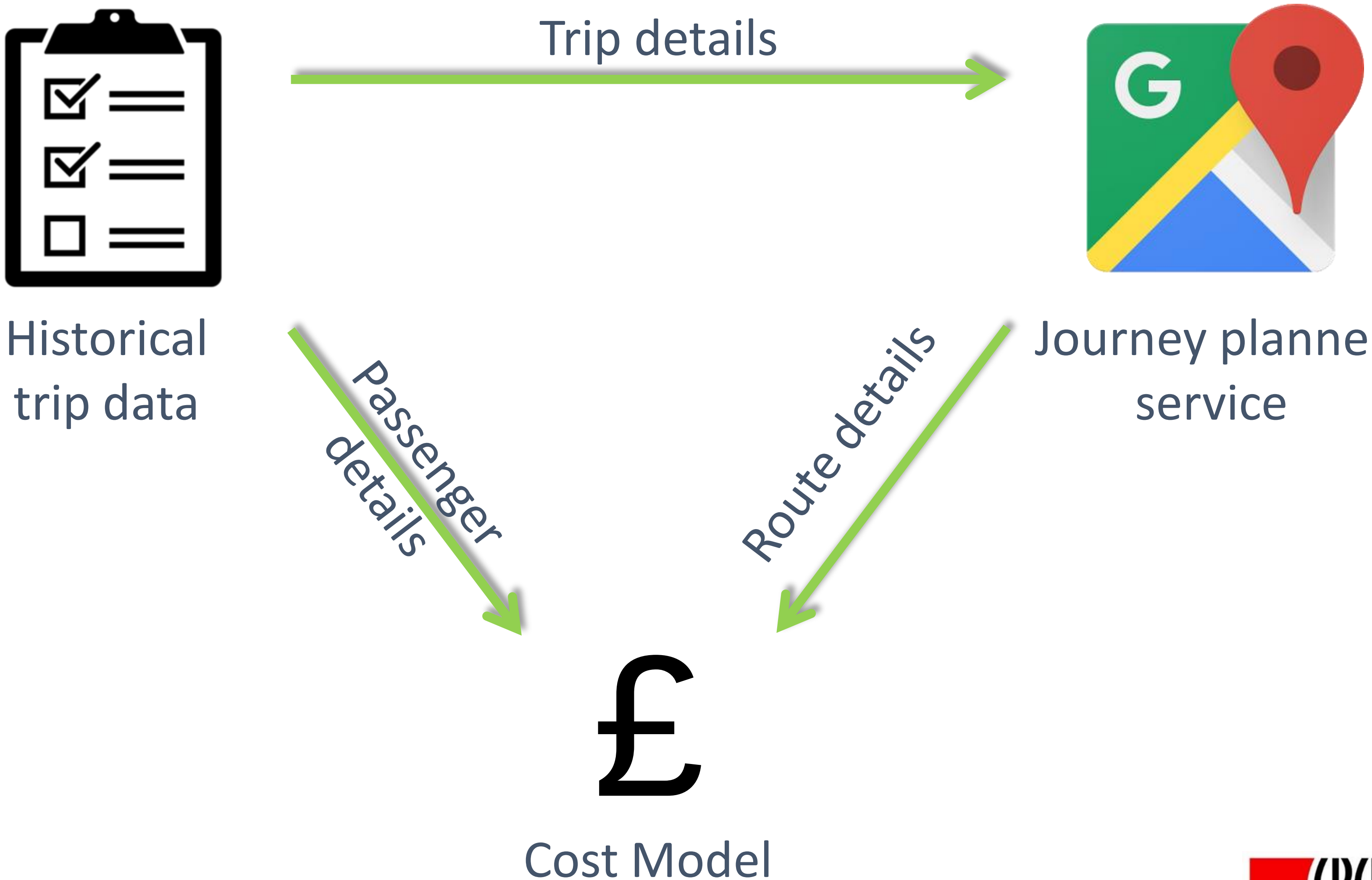
1/49 studies implement sequential optimisation

*Solution: use Bayesian optimisation*

# 3. Optimising hyper-parameters



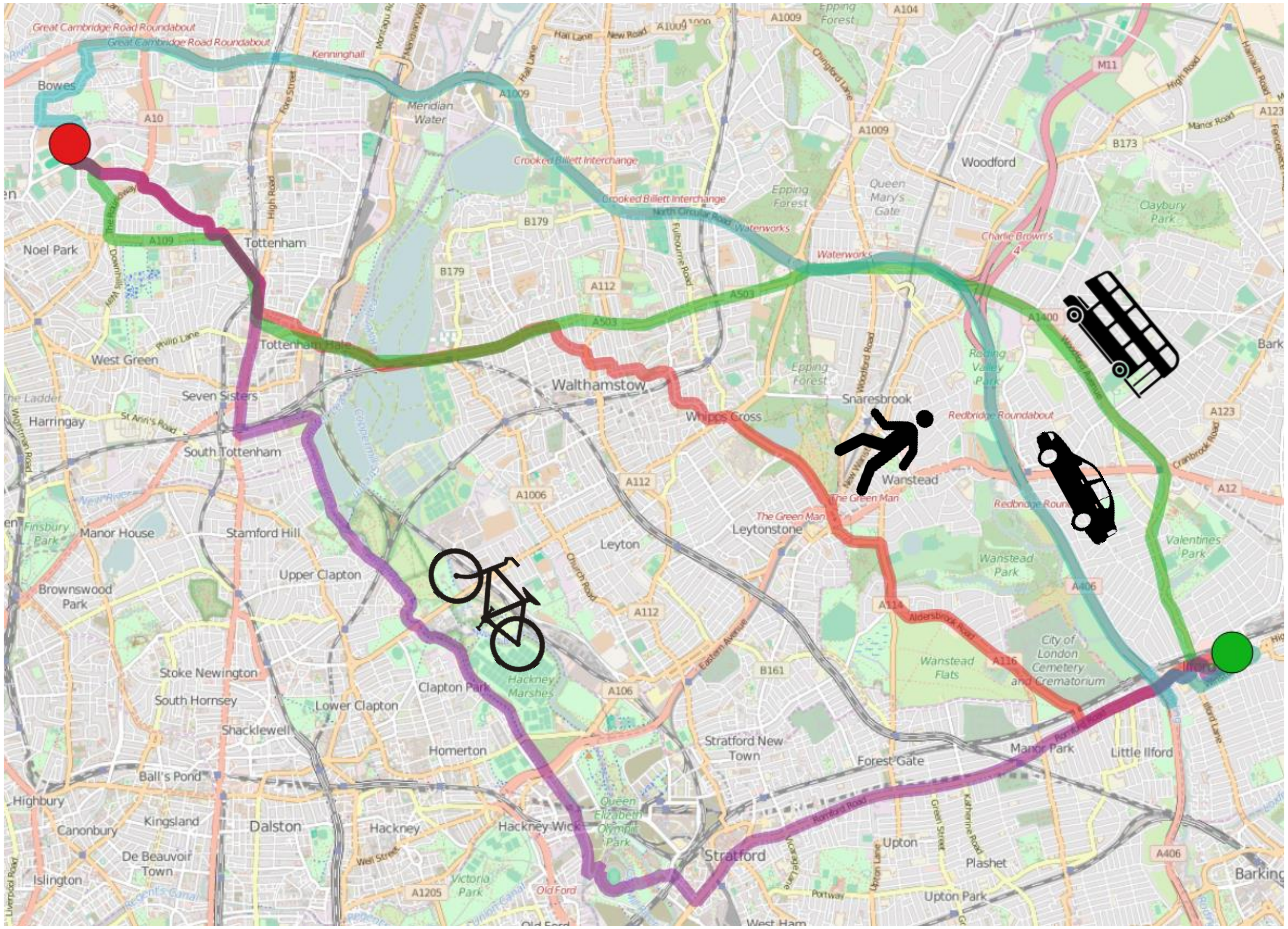
# Dataset




UNIVERSITY OF CAMBRIDGE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Dataset



Bowes
Noel Park
Tottenham
West Green
Seven Sisters
Harringay
South Tottenham
Tottenham Hale
Meridian Water
Walthamstow
Whipps Cross
Snaresbrook
Wanstead
Woodford
Leytonstone
Leyton
Manor House
Stamford Hill
Upper Clapton
Brownswood Park
Stoke Newington
South Hornsey
Lower Clapton
Clapton Park
Hackney Marshes
Shacklewell
Homerton
Ball's Pond
Highbury
Canonbury
Kingsland
Dalston
Hackney
De Beauvoir Town
Islington
Hackney Wick
Stratford
Stratford New Town
Forest Gate
Manor Park
Little Ilford
Upton
Plashet
Upton Park
Barking
Wanstead Flats
Wanstead Park
City of London Cemetery and Crematorium
Valentines Park
Claybury Park
Queen Mary's Gate
Epping Forest

# Methodology

3 years

2012/13-2013/14 train

2014/15 test

10 fold CV to estimate performance on train

# Methodology

| Model |
|---|
| Artificial Neural Network |
| Extra Trees |
| Extreme Gradient Boosting |
| K-Nearest Neighbours |
| Logistic Regression |
| Multinomial Naïve Bayes |
| Random Forest |
| Support Vector Classifier |

# Results – estimate performance

| Model | Trip wise |
|---|---|
| Artificial Neural Network | 0.6739 |
| Extra Trees | 0.4782 |
| Extreme Gradient Boosting | 0.4665 |
| K-Nearest Neighbours | 0.7403 |
| Logistic Regression | 0.6770 |
| Multinomial Naïve Bayes | 0.8626 |
| Random Forest | 0.4930 |
| Support Vector Classifier | 0.6248 |

Cross entropy loss
(Lower is better)

# Results – estimate performance

| **Model** | **Trip wise** | **Household wise** |
|---|---|---|
| Artificial Neural Network | 0.6739 | 0.6783 |
| Extra Trees | 0.4782 | 0.6571 |
| Extreme Gradient Boosting | 0.4665 | **0.6338** |
| K-Nearest Neighbours | 0.7403 | 0.9713 |
| Logistic Regression | 0.6770 | 0.6795 |
| Multinomial Naïve Bayes | 0.8626 | 0.8638 |
| Random Forest | 0.4930 | 0.6578 |
| Support Vector Classifier | 0.6248 | 0.6571 |

# Methodology

Optimise model using 10-fold CV on train

Using trip-wise and household-wise sampling

# Results – test optimised model

| **Model** | **Trip wise** | **Household wise** |
|---|---|---|
| Artificial Neural Network | 0.7018 | 0.6941 |
| Extra Trees | 0.6961 | 0.6791 |
| Extreme Gradient Boosting | 0.7259 | **0.6510** |
| K-Nearest Neighbours | 1.0171 | 1.0397 |
| Logistic Regression | 0.6935 | 0.6930 |
| Multinomial Naïve Bayes | 0.8793 | 0.8766 |
| Random Forest | 0.6832 | 0.6774 |
| Support Vector Classifier | 0.7316 | 0.6702 |

# Results – estimate performance

| **Model** | **Trip wise** | **Household wise** |
|---|---|---|
| Artificial Neural Network | 0.6739 | 0.6783 |
| Extra Trees | 0.4782 | 0.6571 |
| Extreme Gradient Boosting | 0.4665 | **0.6338** |
| K-Nearest Neighbours | 0.7403 | 0.9713 |
| Logistic Regression | 0.6770 | 0.6795 |
| Multinomial Naïve Bayes | 0.8626 | 0.8638 |
| Random Forest | 0.4930 | 0.6578 |
| Support Vector Classifier | 0.6248 | 0.6571 |

# Summary

1. Treating choice prediction as deterministic
2. Using trip-wise sampling with hierarchical data
3. Not optimising model hyper-parameters

# Summary

1. Treating choice prediction as deterministic

2. Using trip-wise sampling with hierarchical data

3. Not optimising model hyper-parameters

# Thank you

**Hillel, Tim, Mohammed Z E B Elshafie, and Ying Jin (2018). "Recreating Passenger Mode Choice-Sets for Transport Simulation: A Case Study of London, UK". In: *Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction* 171.1, pp. 29–42.**

***th389@cam.ac.uk***