15th World Conference on Transport Research
26 – 31 May 2019 Mumbai, India

Hybrid Data Envelopment Analysis for Large-Scale Smartphone Data Modeling

Dimitrios I. Tselentis, PhD, Eleni I. Vlahogianni, PhD, George Yannis, PhD, Loukas Kavouras

Department of Transportation Planning and Engineering,
National Technical University of Athens, Athens, Greece

Scope

- **Methodological approach** for driving safety efficiency **benchmarking**
  - on a trip basis
  - optimization techniques
- **Smartphone** devices
  - large-scale data
  - naturalistic driving conditions
- Estimation time is **exponentially increased**
  - as more driving data are aggregated
  - improve the existing approach

Background (1/2)

- Linear programming for efficiency measurement
  - Data Envelopment Analysis (DEA)
    - Banks, Companies, Hospitals, Staff etc.
    - Transport (Systems, Traffic safety etc.)
  - Driving behaviour research

- Driving behaviour characteristics
  - Speeding
  - Harsh braking/ acceleration/ cornering
  - Seatbelt use
  - Mobile phone use

Background (2/2)

- **Data collection schemes**
  - Smartphones
  - In-vehicle devices
  - On-board diagnostic devices (e.g. OBD-II)

- **Data sources**
  - Naturalistic driving experiments
  - Driving simulator experiments
  - In-depth accident investigation

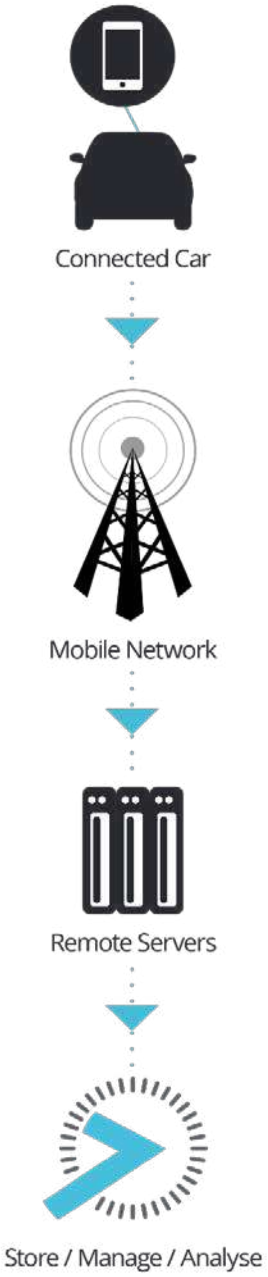
Smartphone data collection (1/2)

- A mobile application to record user's driving behaviour (automatic start / stop)

- A variety of APIs is used to read mobile phone sensor data

- Data is transmitted from the mobile App to the central database

- Data are stored in a sophisticated database where they are managed and processed

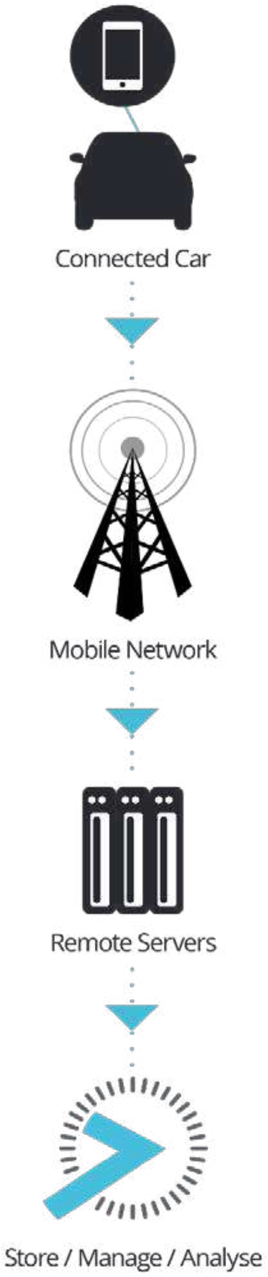


Smartphone data collection (2/2)

- Indicators are designed using:
  - machine learning algorithms
  - big data mining techniques

- Data collected for 10,088 trips from 88 drivers (From July to December 2016)

- The database analyzed was in .csv format
  - Drivers’ trips are stored per row, the characteristics of which are stored in each column’s variables



Data envelopment analysis (1/2)

- Optimization technique
  - Performance measurement using DEA
  - Companies, banks, hospitals, staff etc.

- Significant computation cost
  - Exact solution
  - Reduced Basis Entry (RBE)
  - Convex Hull (CH)
  - Reduce processing time

- Decision-Making-Unit (DMU)
  - Factory, company etc.
  - Trips, drivers
    - variables are continuous and quantitative
    - a driver should reduce the frequency of his driving characteristics for a given mileage

15th World Conference on Transport Research
26 – 31 May 2019 Mumbai, India

Data envelopment analysis (2/2)

- **Efficiency index** Driving_efficiencyB

- **Input-oriented** DEA
  - minimize inputs (number of HA, HB events etc.) per driving distance

- Constant-returns-to-scale (**CRS**) problem
  - the sum of all inputs changes proportionally to the sum of driving output (distance)

- **Efficient level** of driving characteristics for a trip/ driver
  - inputs
  - outputs

$$\min(\text{Driving\_Efficiency}_{\text{B}})$$

Subject to the following constraints:

$$\text{Driving\_Efficiency}_{\text{B}} * x_o - X * \lambda \geq 0$$

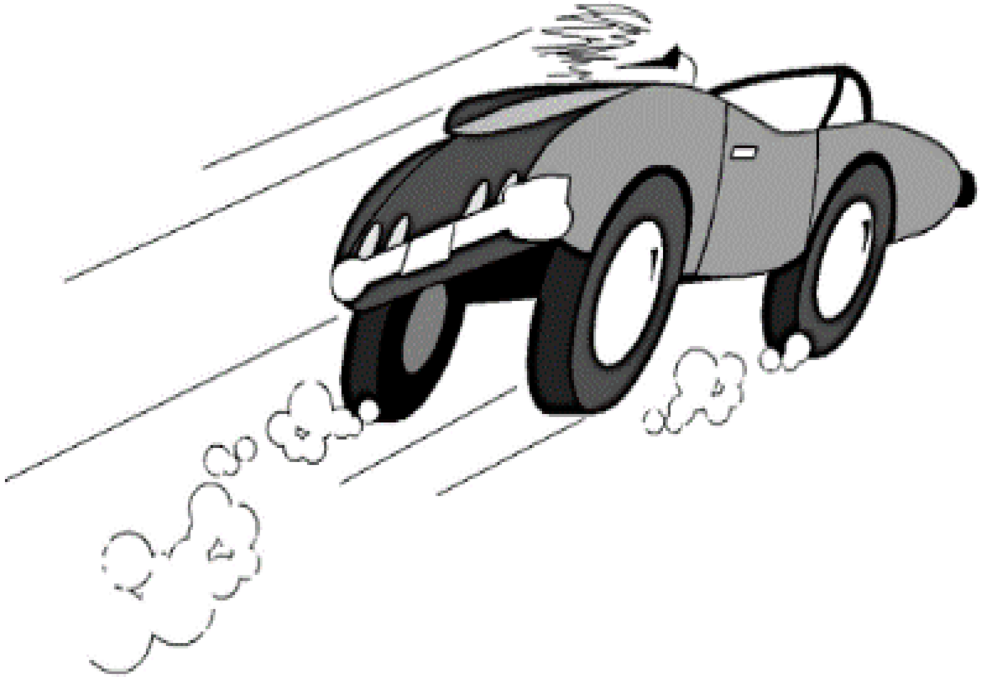
$$\text{Y} * \lambda \geq y_o$$

$$\lambda_i \geq 0 \forall \lambda_i \in \lambda$$

$$Metric_i = \sum_{j=1}^{m} \lambda_j * Metric_j$$

$$distance_{urban} = distance_i \,/\, \text{Driving\_Efficiency}_i$$

Efficiency index parameters



Risk exposure indicators:

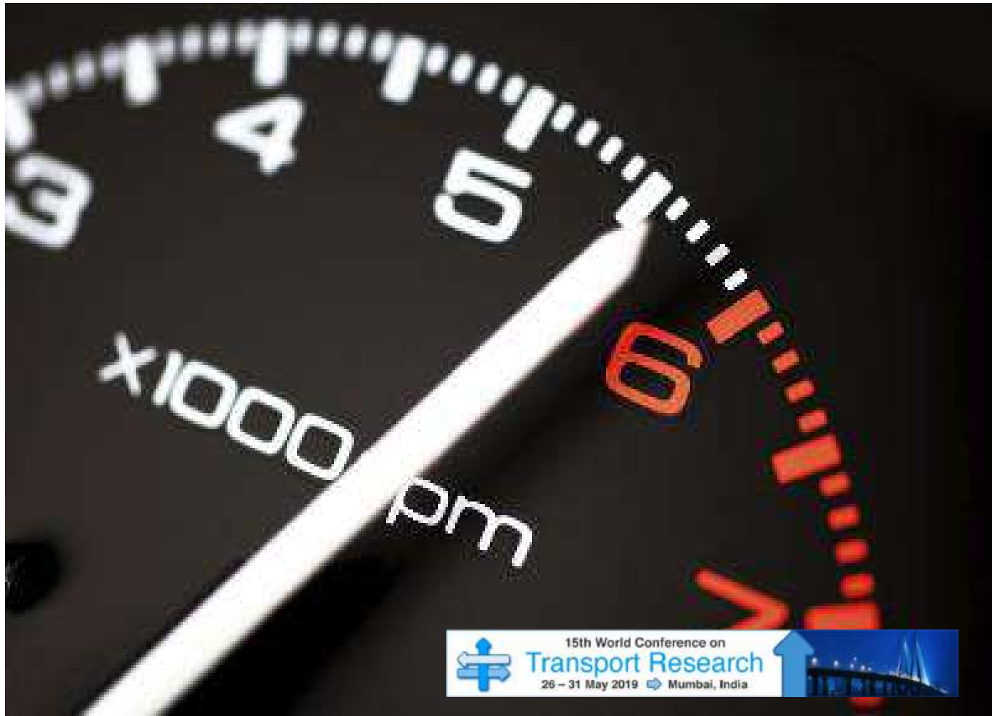
- Total distance travelled

Driving behaviour indicators:

- Harsh events
  - Number of harsh braking (longitudinal acceleration) (HA)
  - Number of harsh acceleration (longitudinal acceleration) (HB)
- Speeding (SP)
  - seconds driving over the speed limit
- Mobile phone use distraction (MU)
  - seconds using the mobile phone

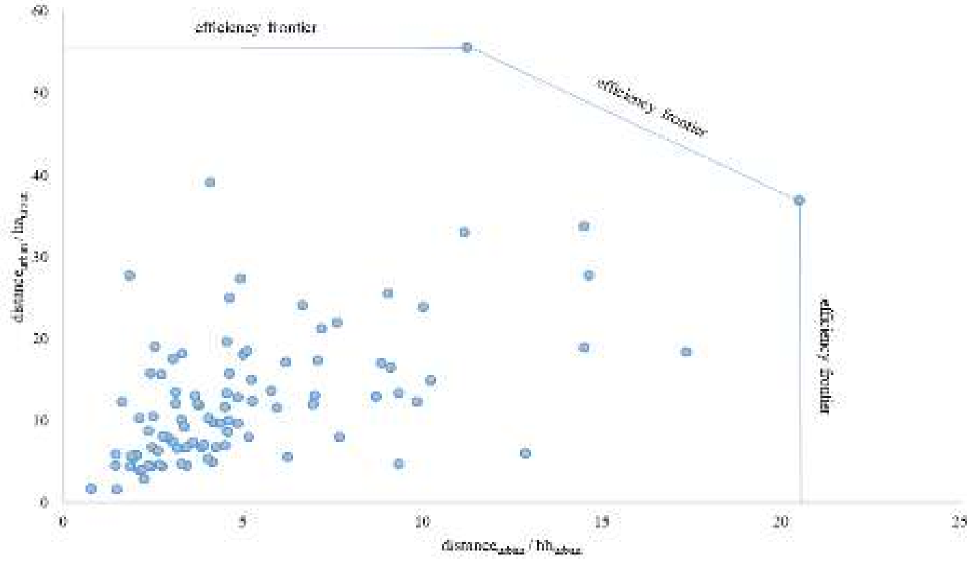
Road types:

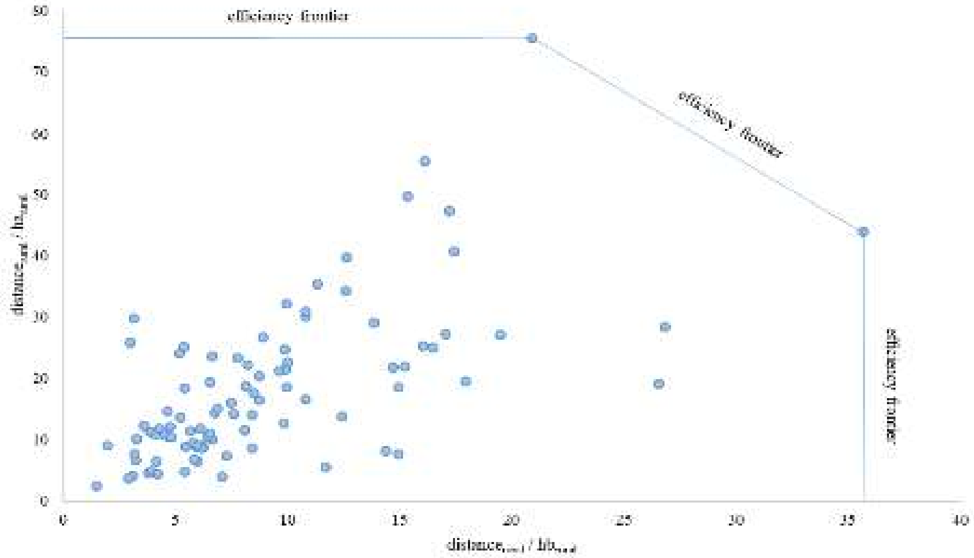
- Urban road network (signalized or not network, speed limit ≤ 50 km/h)
- Urban express road network (speed limit 50 – 90 km/h)
- Highways (speed limit ≥ 90 km/h)



Driving efficiency analysis using DEA

- DEA results **2-D illustration**
  - Urban
  - Rural
- DEA **inputs**
  - Number of harsh acceleration events
  - Number of harsh braking events
- DEA **outputs**
  - Trip distance





Data preparation

- Data are anonymized
  - user-agnostic approach
  - identify driving behaviors and patterns
  - causality between behaviour and other factors
  - large-scale samples
  - no information on demographics or accident record

- Python programming language
  - filter aggregate data
  - retain only necessary information
  - aggregate data
  - data analysis

Trip efficiency analysis (1/3)

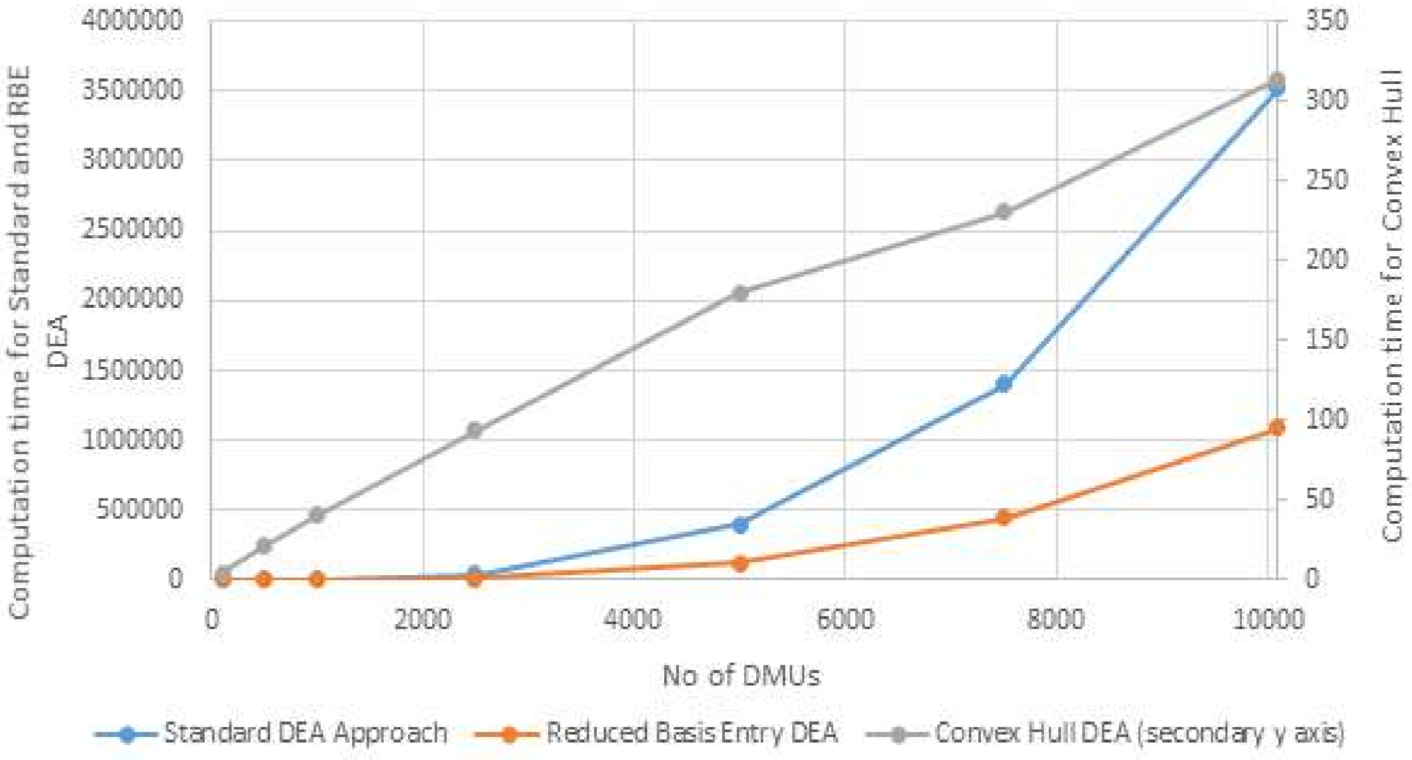
- **Convex Hull** technique outperforms
- Significant **time reduction**
  - As the database becomes larger
  - Convex Hull DEA – 5 minutes
  - RBE DEA – 12.6 days
  - Standard DEA – 40.7 days

	Computation time (sec)			CH DEA % computation time improvement over		RBE DEA % computation time improvement over Standard DEA
No of DMUs	Standard DEA Approach	RBE DEA	Convex Hull DEA	Standard DEA Approach	RBE DEA	
100	11	6	4	63.64%	33.33%	45.45%
500	477	169	21	95.60%	87.57%	64.57%
1000	3250	1121	41	98.74%	96.34%	65.51%
2500	44435	15570	94	99.79%	99.40%	64.96%
5000	398485	123986	180	99.95%	99.85%	68.89%
7500	1400909	444498	231	99.98%	99.95%	68.27%
10088	3519372	1089731	314	99.99%	99.97%	69.04%

*** Inputs = [ha_{urban}, ha_{rural}, $ha_{highway}$], Outputs = [$distance_{urban}$, $distance_{rural}$, $distance_{highway}$]**

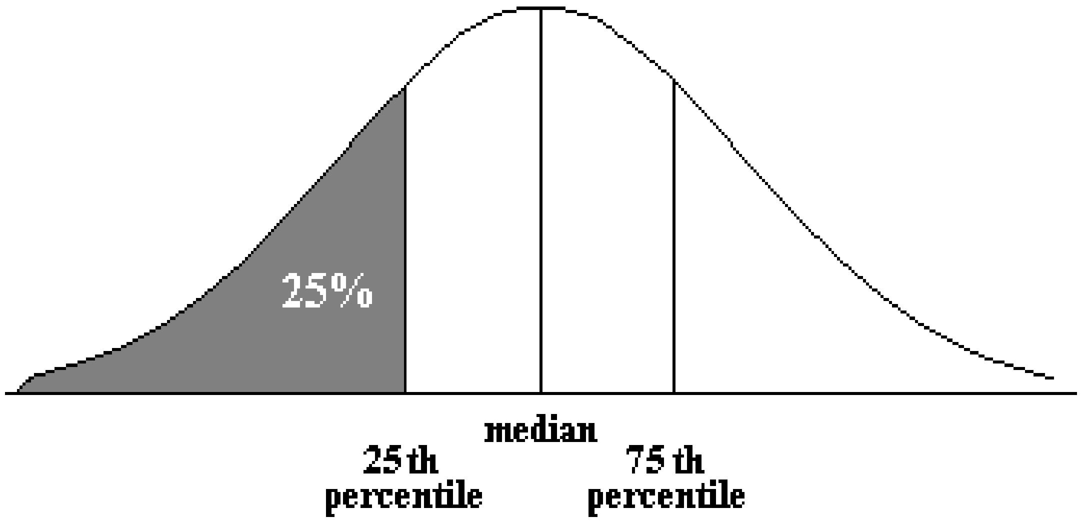
Trip efficiency analysis (2/3)

- Convex Hull DEA
  - linearly increased
- Standard and RBE DEA
  - Exponentially increased



Trip efficiency analysis (3/3)

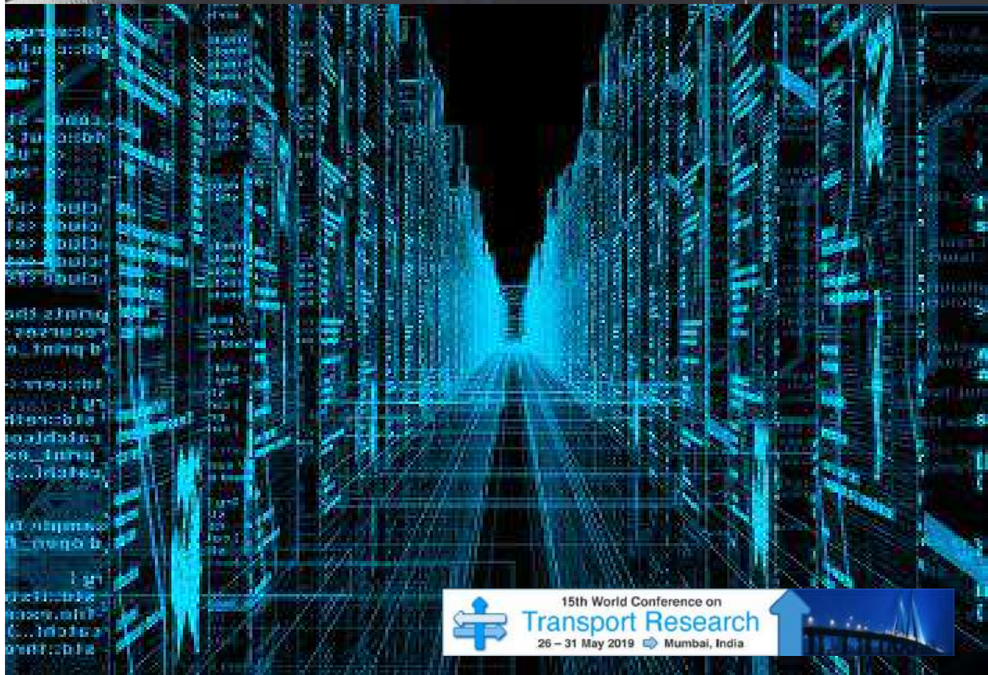
- Least efficient trips
- Efficiency estimation
- Sort
  - larger to smaller
- Percentile
  - 5%, 10%, 25% etc.



Conclusions

- **Innovative** methodological approach for driving efficiency benchmarking as well as for estimating the efficient level of metrics for each trip.

- The integration of DEA with the convex hull algorithmic approach yielded **significantly better** results than the rest of the approaches tested.



Future research

- Exploit a larger driving sample
- Overcome DEA limitations – zero sum input attributes
- Increase the number of attributes
  - headways
  - lane changing
  - eye movement
  - drowsiness

15th World Conference on
Transport Research
26 – 31 May 2019 Mumbai, India

Hybrid Data Envelopment Analysis for Large-Scale Smartphone Data Modeling

Dimitrios I. Tselentis, PhD, Eleni I. Vlahogianni, PhD, George Yannis, PhD, Loukas Kavouras

Department of Transportation Planning and Engineering, National Technical University of Athens, Athens, Greece