Using mobile network big data for land use classification

Kaushalya Madhawa, Sriganesh Lokanathan, Danaja Maldeniya, Rohan Samarajiva

CPR*south* 2015

Taipei City, Taiwan

25th August 2015

LIRNE*asia*

Pro-poor. Pro-market.

This work was carried out with the aid of a grant from the International Development Research Centre, Canada and the Department for International Development UK..

Implications of using mobile network big data for urban policy

- Almost real-time monitoring of urban land use
- Help align master plan to reality
- Complement infrequent and expensive surveys

The data: historical and anonymized Call Detail Records (CDRs) from Sri Lanka

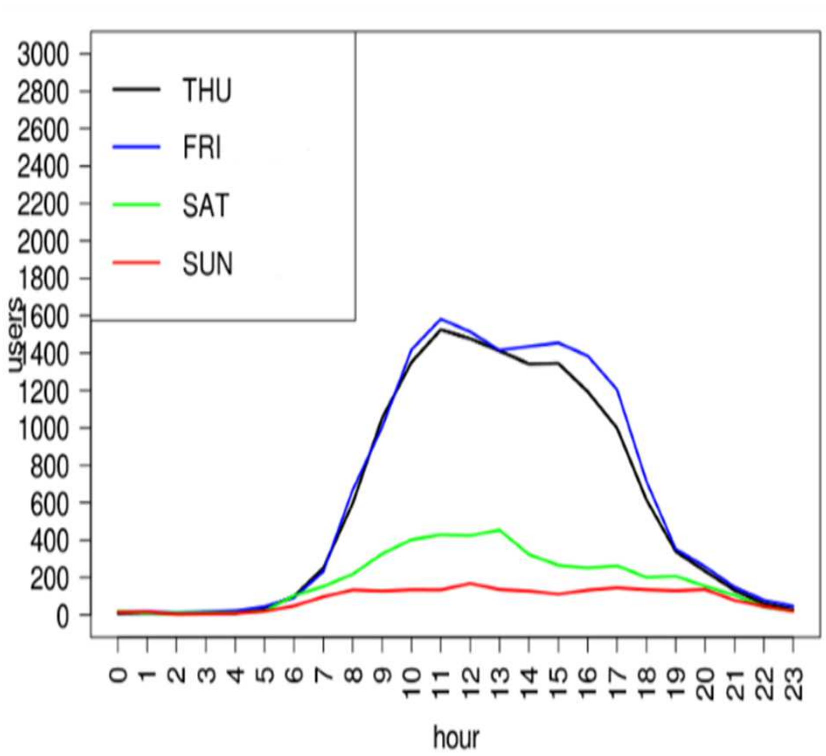
- Call Detail Record (CDR):
  - Records of all calls made and received by a person created mainly for the purposes of billing
  - Similar records exist for all SMS-es sent and received as well as for all Internet sessions

Calling Party Number	Called Party Number	Caller Cell ID	Call Time	Call Duration
A24BC1571X	B321SG141X	3134	13-04-2013 17:42:14	00:03:35

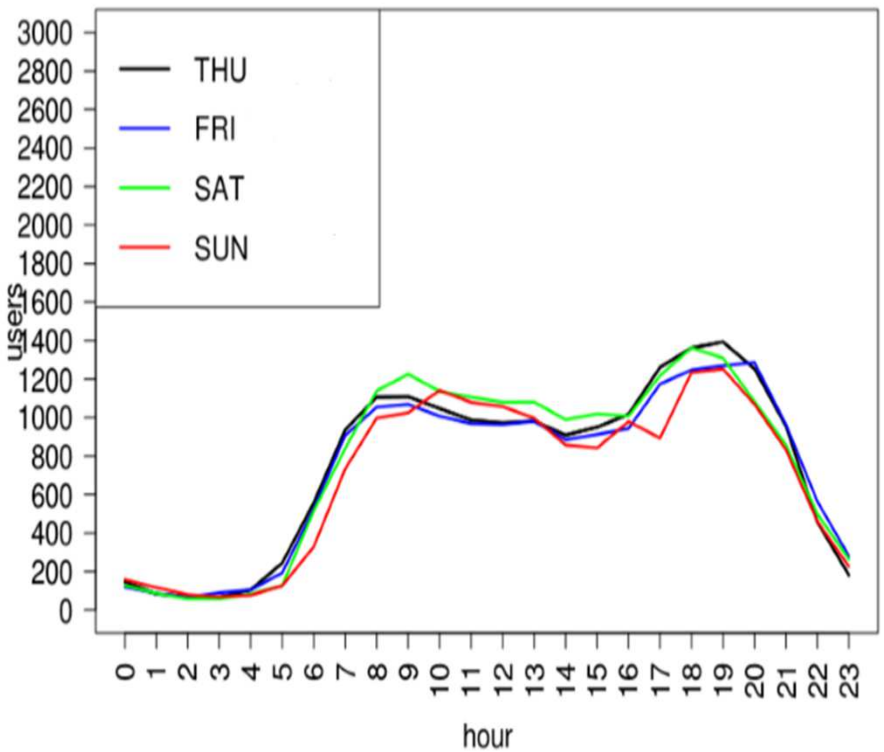
  - The Cell ID in turn has a lat-long position associated with it
- CDR data for 1 month in 2013
  - Covers under 10 million SIMs
  - Nearly 1.5 billion records of calls made and received

LIRNEasia
Pro-poor. Pro-market.

The hourly loading of base stations reveals distinct patterns



Type X: ?



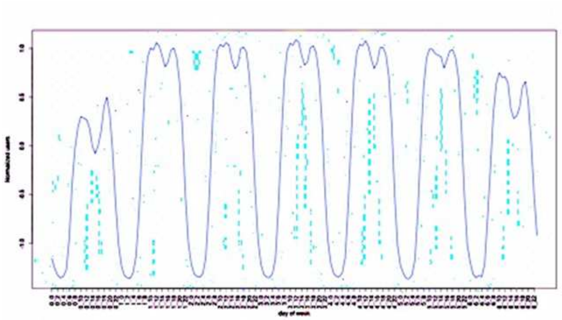
Type Y: ?

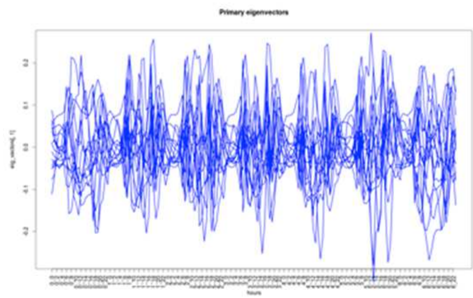
- We can use this insight to group base stations into different groups, using unsupervised machine learning techniques

LIRNEasia
Pro-poor. Pro-market.

Methodology

- The time series of users connected at a base station contains variations, that can be grouped by similar characteristics
- A month of data is collapsed into an indicative week (Sunday to Saturday), with the time series normalized by the z-score
- Principal Component Analysis(PCA) is used to identify the discriminant patterns from noisy time series data

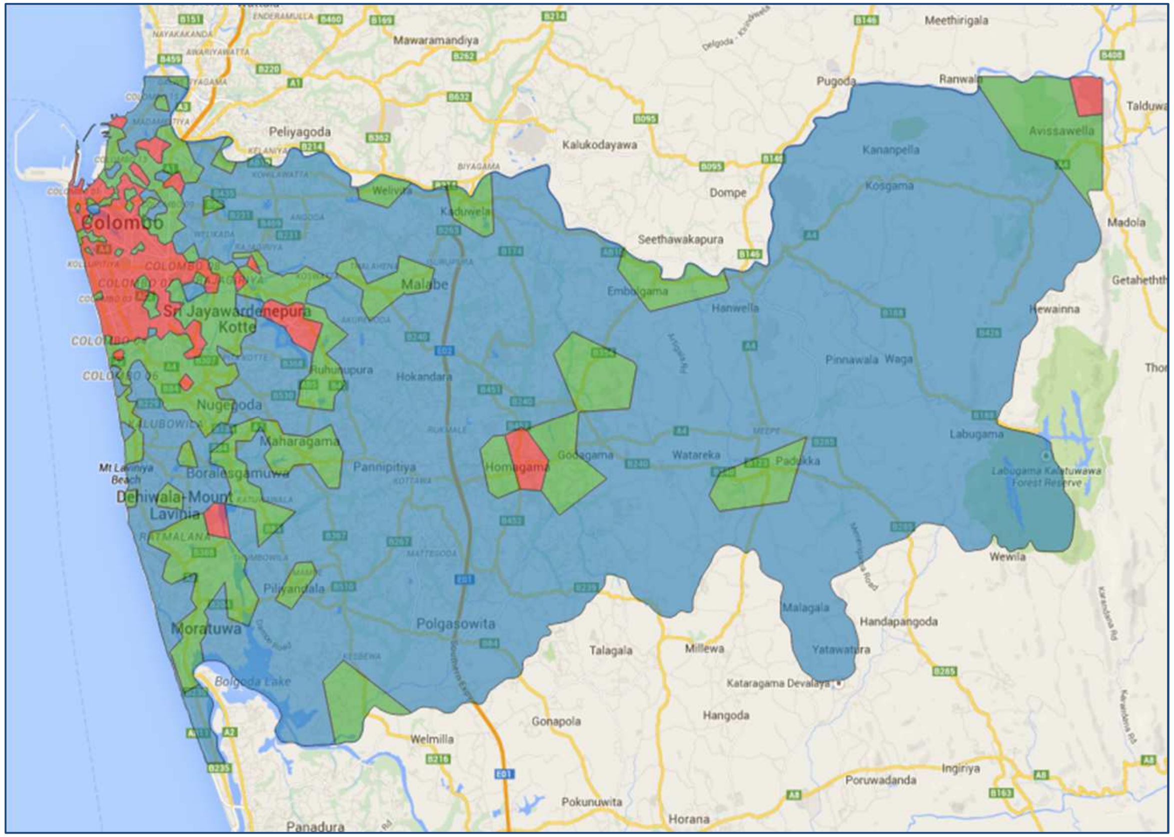




  - Each base station's pattern is filtered into 15 principal components (covering 95% of the data for that base station)
- Using the 15 principal components, we cluster all the base stations into 3 clusters in an unsupervised manner using k-means algorithm

LIRNEasia
Pro-poor. Pro-market.

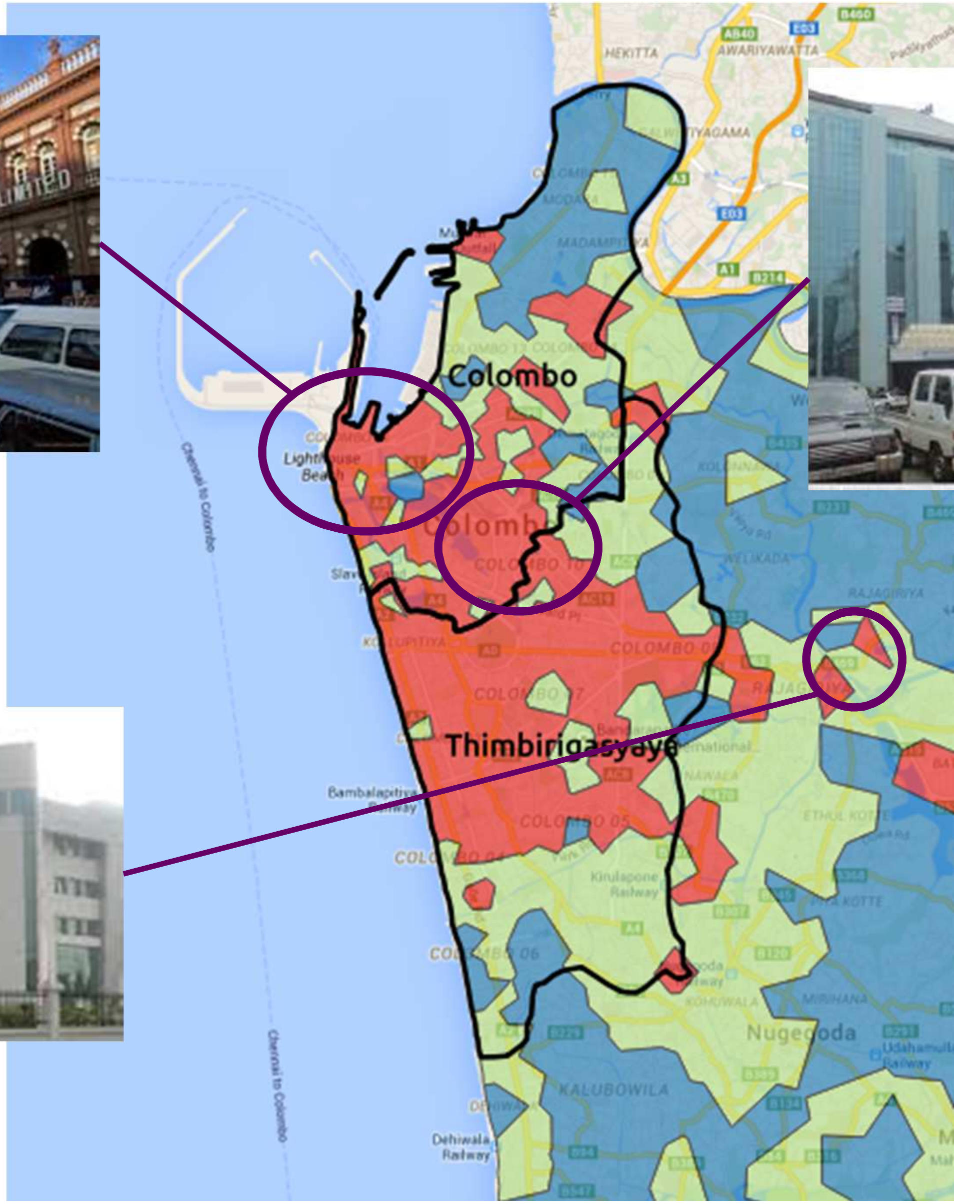
Three spatial clusters in Colombo District



- Cluster-1 exhibits patterns consistent with commercial area
- Cluster-3 exhibits patterns consistent with residential area
- Cluster-2 exhibits patterns more consistent with mixed-use

LIRNEasia
Pro-poor. Pro-market.

Our results show Central Business District (CBD) in Colombo city has expanded



LIRNEasia
Pro-poor. Pro-market.

Small area in NE corner of Colombo District classified as belonging to Cluster 1?

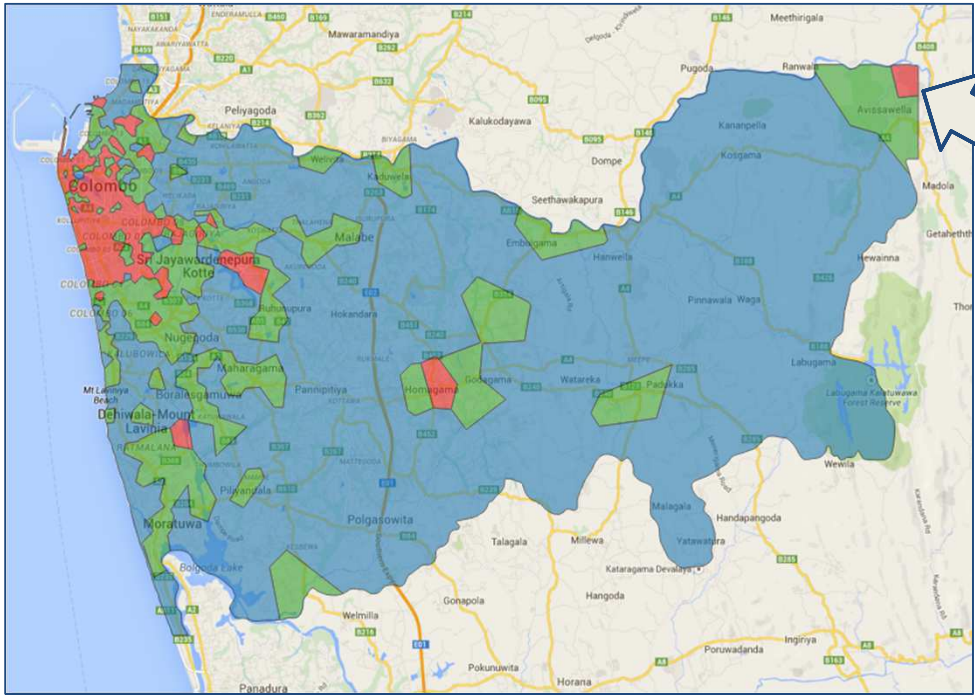


Photo ©Senanayaka Bandara - [Panoramio]()

Seethawaka Export Processing Zone

LIRNEasia Pro-poor. Pro-market.

We use silhouette coefficients to understand the quality of the clustering

- Silhouette coefficient indicates quality of clustering

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

a(i) - average distance of i with all other data within the same cluster

b(i) - average distance of i with all other data within the neighboring cluster

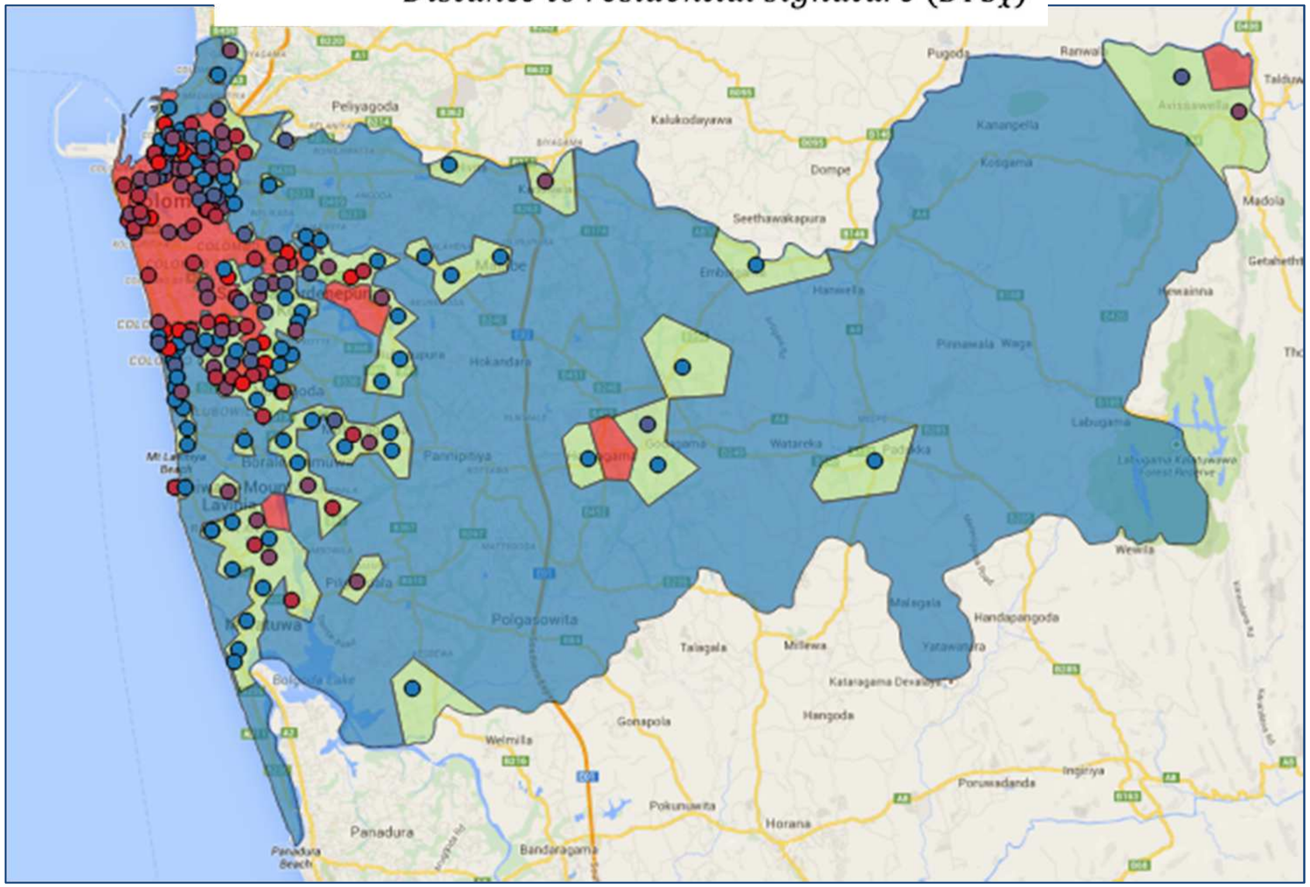
- Based on the s-values, Cluster 3 is the least coherent amongst the three

Cluster	Avg. Silhouette Coefficient
1 – Commercial	0.46
2 – Residential	0.36
3 – Mixed-use	0.22

Internal variations in mixed use regions: More commercial or more residential?

- To evaluate the relative closeness to the other two clusters, we define extent of commercialization as:

$$C(BTS_x) = \frac{Distance\ to\ commercial\ signature\ (BTS_x)}{Distance\ to\ residential\ signature\ (BTS_x)}$$

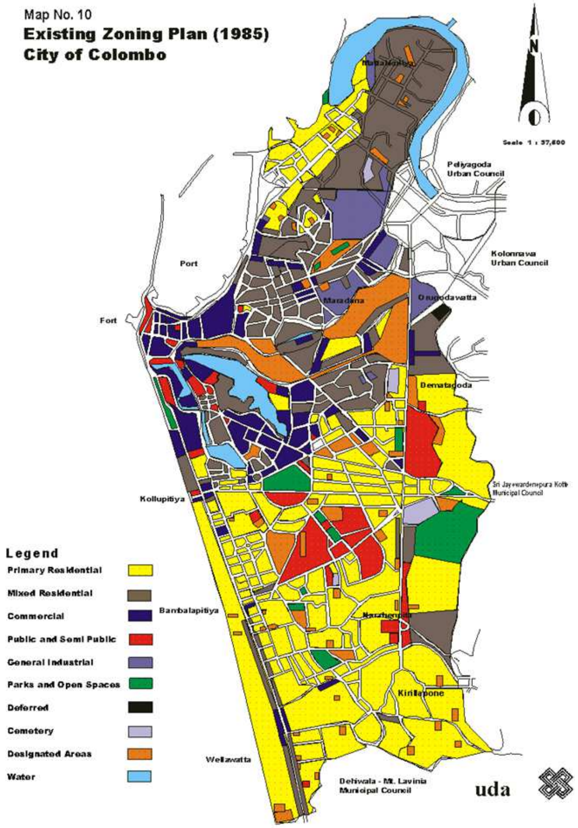


Blue dots: more residential than commercial

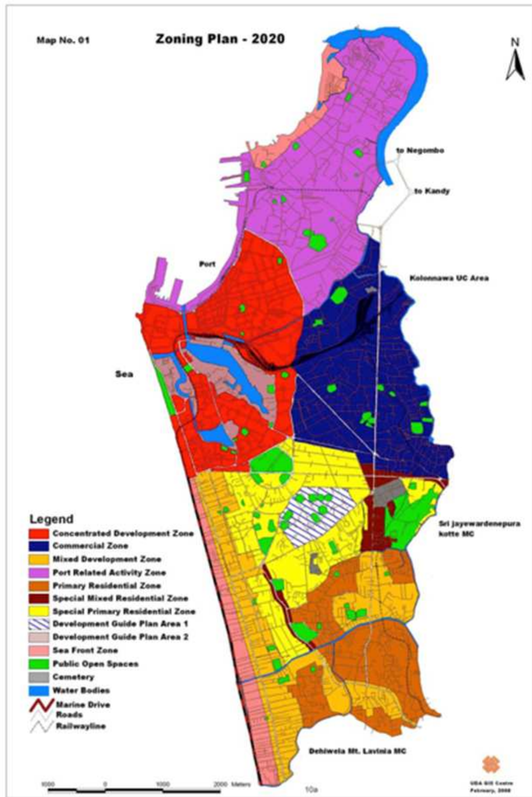
Red dots: more commercial than residential

Plans & reality

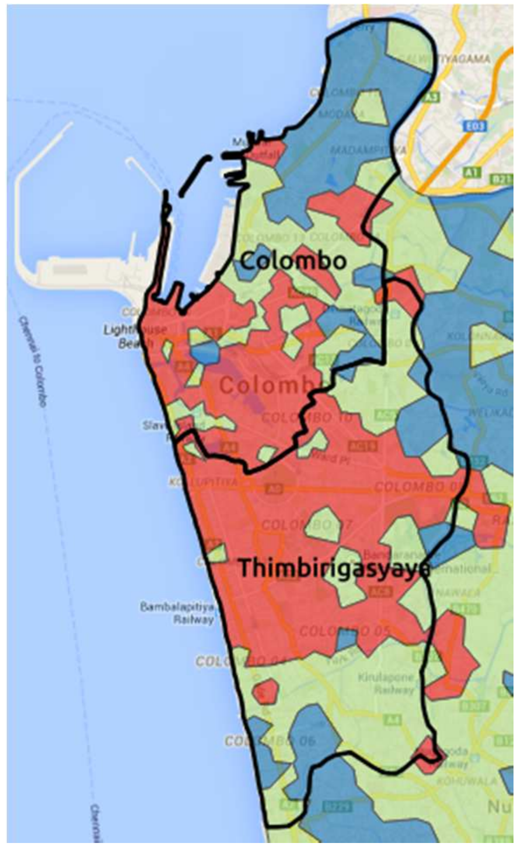
1985 Plan



2020 UDA Plan



2013 reality



Policy implications and future work

- Almost real-time monitoring of urban land use
  - We are currently working on understanding finer temporal variations in zone characteristics (especially the mixed-use areas)
- Help align master plan to reality
- Can complement infrequent & expensive surveys
- LIRNE*asia* is working to unpack the identified categories further, e.g.,
  - Entertainment zones that show evening activity