Planning meeting: Forecasting propagation of dengue/zika in Sri Lanka with Mobile Network Big Data

LIRNE*asia*

http://lirneasia.net/projects/bd4d/

6th May 2016

LIRNEasia

Pro-poor. Pro-market

IDRC CRDI Canada

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..

UKaid from the British people

Objective of this meeting

- Discuss prior work in using mobile network big data for building models of infectious disease (dengue) propagation
  - Pakistan case study
  - Preliminary work in Sri Lanka
- Collaboratively brainstorm research parameters

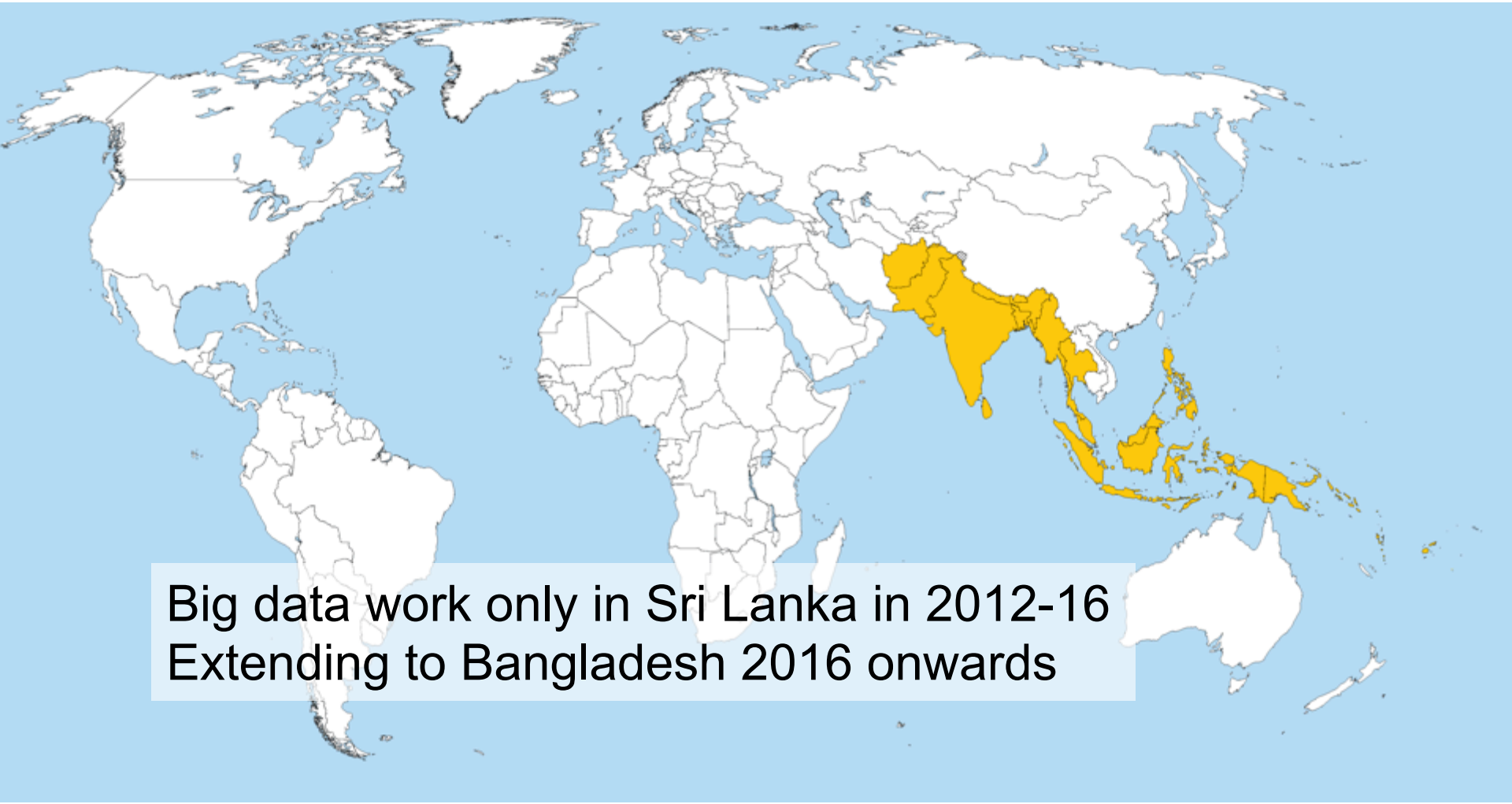
Our mission

Catalyzing policy change through research to improve people's lives in the emerging Asia Pacific by facilitating their use of hard and soft infrastructures through the use of knowledge, information and technology.

LIRNEasia

Pro-poor. Pro-market.

Where we work



Big data work only in Sri Lanka in 2012-16
Extending to Bangladesh 2016 onwards

LIRNEasia
Pro-poor. Pro-market.

LIRNE*asia* Data Analytics Advisory Council

- **Dr. Linus Bengtsson, PhD**
  Co-Founder and Executive Director, Flowminder
- **Joshua Blumenstock, PhD**
  Assistant Professor, School of Information
  Director, Data Science and Analytics Laboratory
  University of Washington
- **Nitesh Chawla, PhD**
  Professor of Computer Science and Engineering
  Director of The Interdisciplinary Center for Network Science & Applications (iCeNSA)
  University of Notre Dame
- **Vanessa Frias-Martinez, PhD**
  Assistant Professor, School of Information Studies
  University of Maryland
- **Amal Kumarage, PhD**
  Senior Professor at Department of Transport & Logistics Management, University of Moratuwa
- **Ashwin Mahesh, PhD**
  Founder & CEO, Mapunity
- **P.K.S. Mahanama, PhD**
  Professor and Former Dean of Department of Town and Country Planning, University of Moratuwa
- **Wasan Pattara-atikom, PhD**
  Principal Researcher & Head of Intelligent Transportation System Laboratory, NECTEC
- **Amal Shehan Perera, PhD**
  Senior Lecturer, Department of Computer Science and Engineering, University of Moratuwa
- **Louiqa Raschid, PhD**
  Professor, Smith School of Business, Center for Bioinformatics and Computational Biology
  UMIACS and the Department of Computer Science
  Robert H. Smith School of Business, University of Maryland
- **Srinath Perera, PhD**
  Vice President, Research, WSO2 Inc.
- **Prabir Sen, PhD**
  Former Chief Data Scientist
  Infocomm Development Authority of Singapore
- **Hetan Shah**
  Executive Director, Royal Statistical Society
- **Ryosuke Shibasaki, PhD**
  Professor, Dr.Eng. Center for Spatial Information Science
  University of Tokyo
- **Linnet Taylor, PhD**
  Marie Curie Research Fellow, Faculty of Social and Behavioural Sciences, University of Amsterdam
- **Ruvan Weerasinghe, PhD**
  Senior Lecturer, University of Colombo - School of Computing
- **Arj Wignaraja**
  Vice President of Operations, Remote Sensing Metrics, LLC

Big data

- An all-encompassing term for any collection of data sets so large or complex that it becomes difficult to process using traditional data processing applications.
- Challenges include: analysis, capture, curation, search, sharing, storage, transfer, visualization, and privacy violations.
- Examples:
  - 100 million Call Detail Records per day generated by Sri Lanka mobile operators
  - 45 Terabytes of data from Hubble Telescope

Why big data? Why now?

- Proximate causes
  - Increased "datafication": Very large sets of schema-less (unstructured, but processable) data now available
  - Advances in memory technology: No longer is it necessary to archive most data and work with small subset
  - Advances in software: MapReduce, Hadoop

If we want comprehensive coverage of the population, what are the sources of big data in developing economies?

- Administrative data
  - E.g., digitized medical records, insurance records, tax records
- Commercial transactions (transaction-generated data)
  - E.g., Stock exchange data, bank transactions, credit card records, supermarket transactions connected by loyalty card number
- Sensors and tracking devices
  - E.g., road and traffic sensors, climate sensors, equipment & infrastructure sensors, mobile phones communicating with base stations, satellite/ GPS devices
- Online activities/ social media
  - E.g., online search activity, online page views, blogs/ FB/ twitter posts

Currently only mobile network big data has broad population coverage

	Mobile SIMs/100	Internet users/100	Facebook users/100
Myanmar	50	2	12
Bangladesh	76	10	9
Pakistan	73	14	11
India	73	18	9
Sri Lanka	107	26	16
Philippines	112	40	41
Indonesia	125	17	25
Thailand	143	35	49

Source: ITU Measuring Information Society 2015; Facebook advantage portal

LIRNEasia
Pro-poor. Pro-market.

Big Data used in the research

- Multiple mobile operators in Sri Lanka have provided four different types of meta-data
  - Call Detail Records (CDRs)
    - Records of calls
    - SMS
    - Internet access
  - Airtime recharge records
- Data sets do not include any Personally Identifiable Information
  - All phone numbers are pseudonymized
  - LIRNE*asia* does not maintain any mappings of identifiers to original phone numbers
- Cover 50-60% of users; very high coverage in Western (where Colombo the capital city in located) & Northern (most affected by civil conflict) Provinces, based on correlation with census data

How mobile network operators see the world

- **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
94777322912	94112327812	3134	13-04-2013 17:42:14	00:03:35

  - The Cell ID in turn has a lat-long position associated with it.
- **Airtime reload records**
  - Records of all airtime reloads performed by prepaid SIMs
  - Each row corresponds to a record of one person's activity:

Number	Type of recharge	Starting balance	Amount	Time
94777322912	CARD	0.41	50	13-04-2013 17:42:14

LIRNEasia
Pro-poor. Pro-market.

Mobile network big data + other data ➔ rich, timely insights

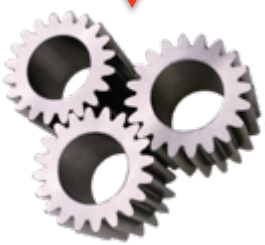
Mobile network big data (CDRs, Internet access usage, airtime recharge records)

Other Data Sources

1. Data from Dept. of Census & Statistics
2. Transportation data
3. Health data
4. Financial data
5. Etc.

Construct Behavioral Variables

1. Mobility variables
2. Social variables
3. Consumption variables



Developmental insights

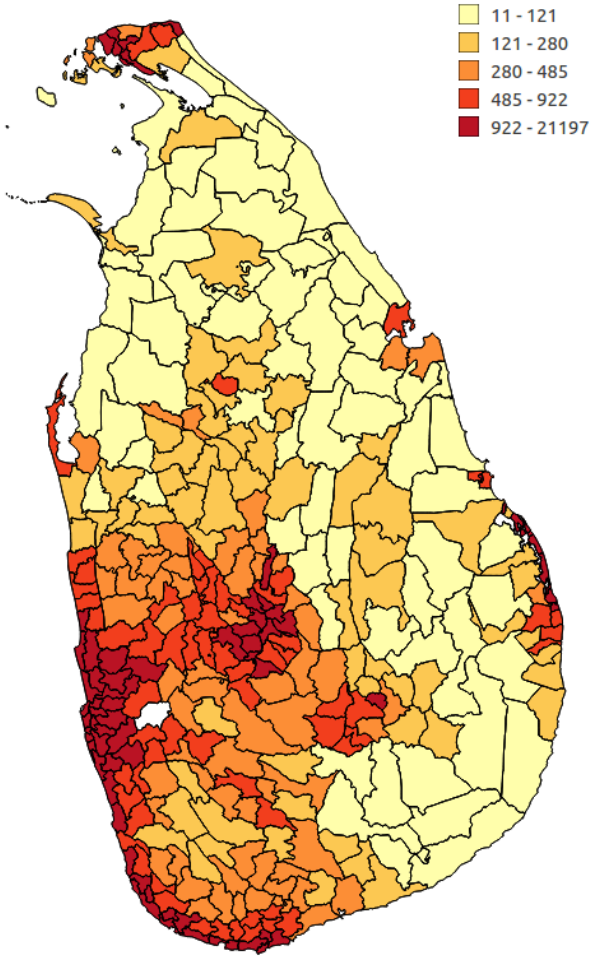
Public purposes

1. Transportation & Urban planning
2. Crises response + DRR
3. Health services
4. Poverty mapping
5. Financial inclusion

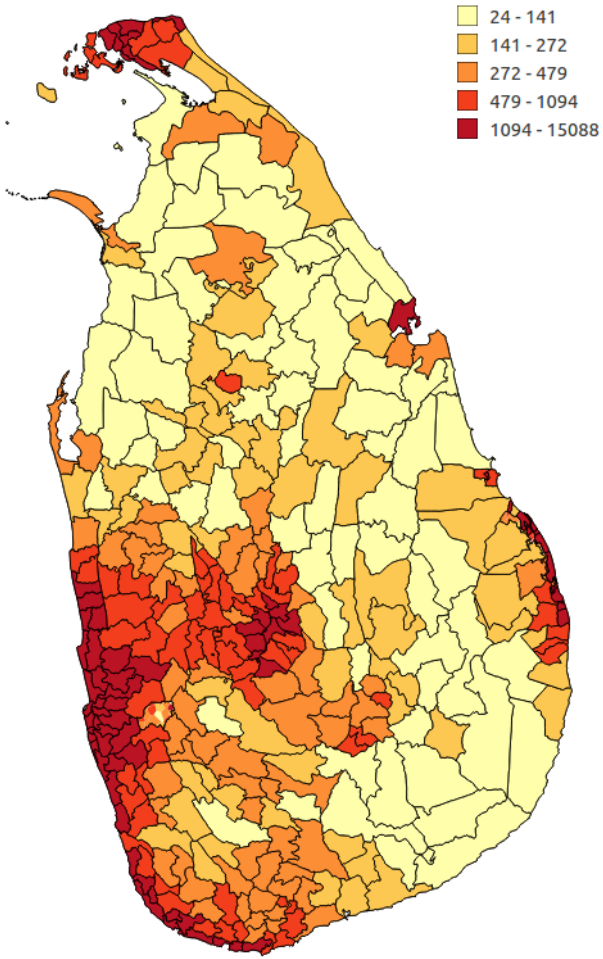
What can you do with such data?

Granular & high-frequency estimates of population density

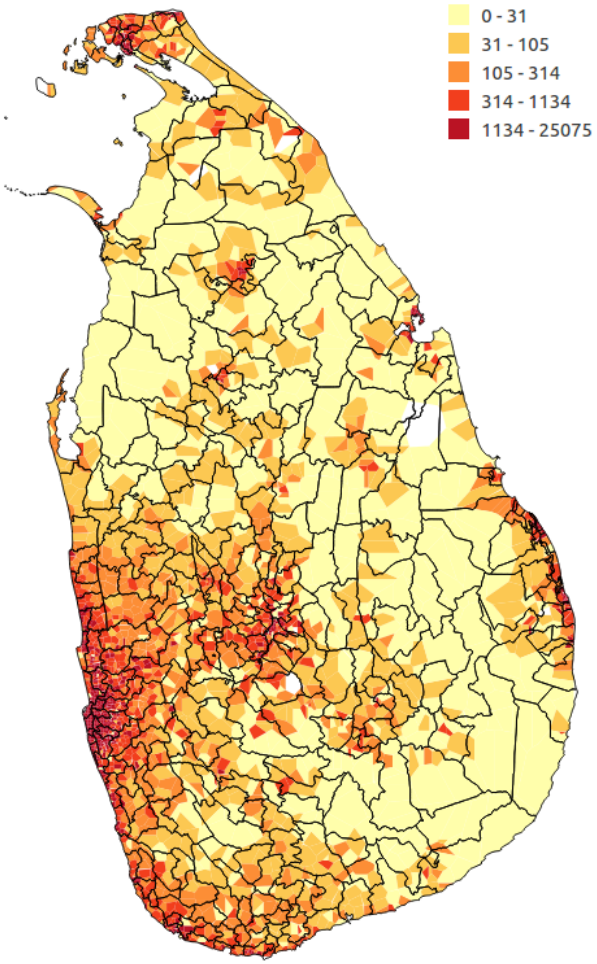
DSD population density from 2012 census



DSD population density estimate from MNBD



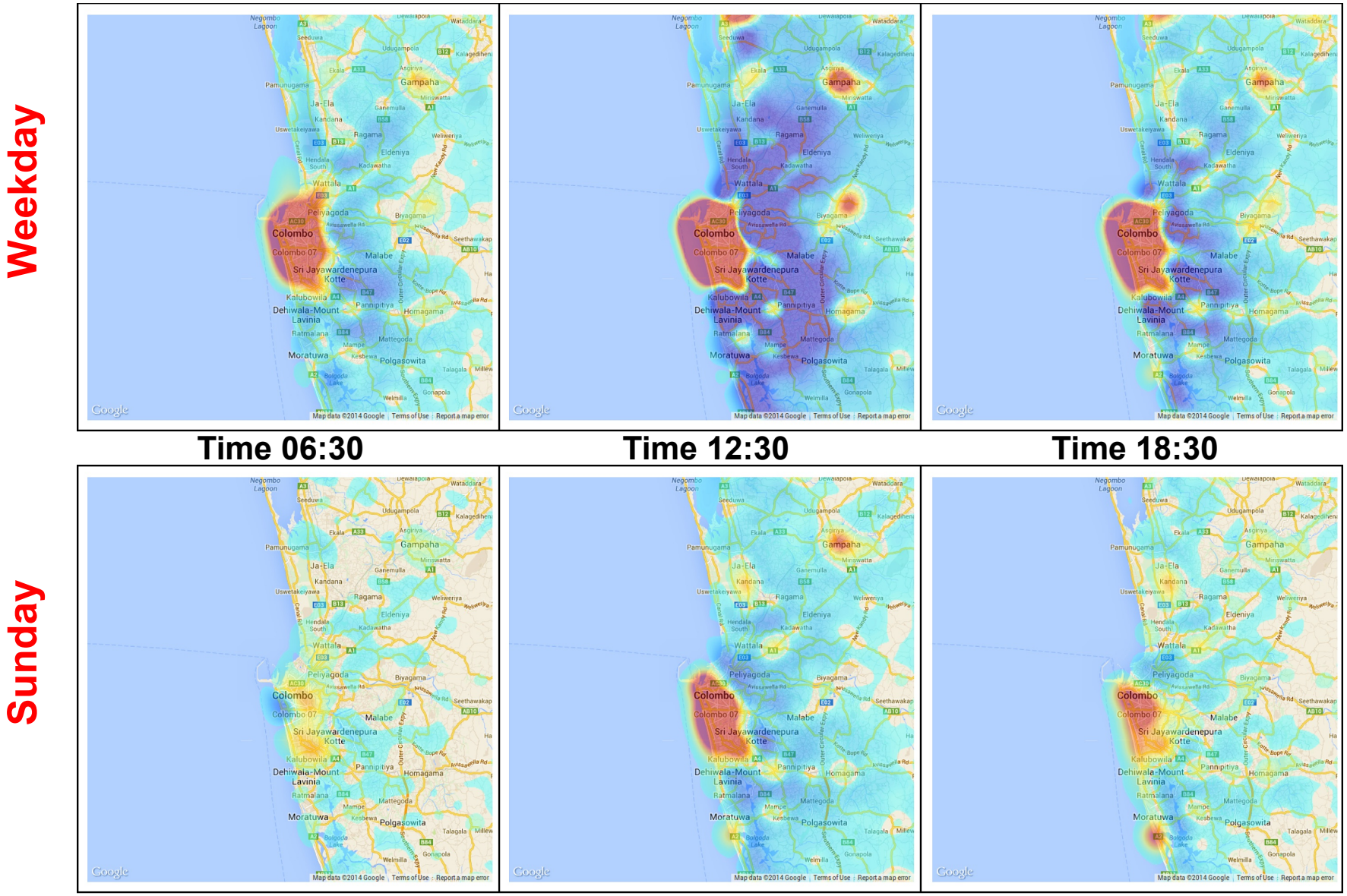
Voronoi cell population density estimate from MNBD



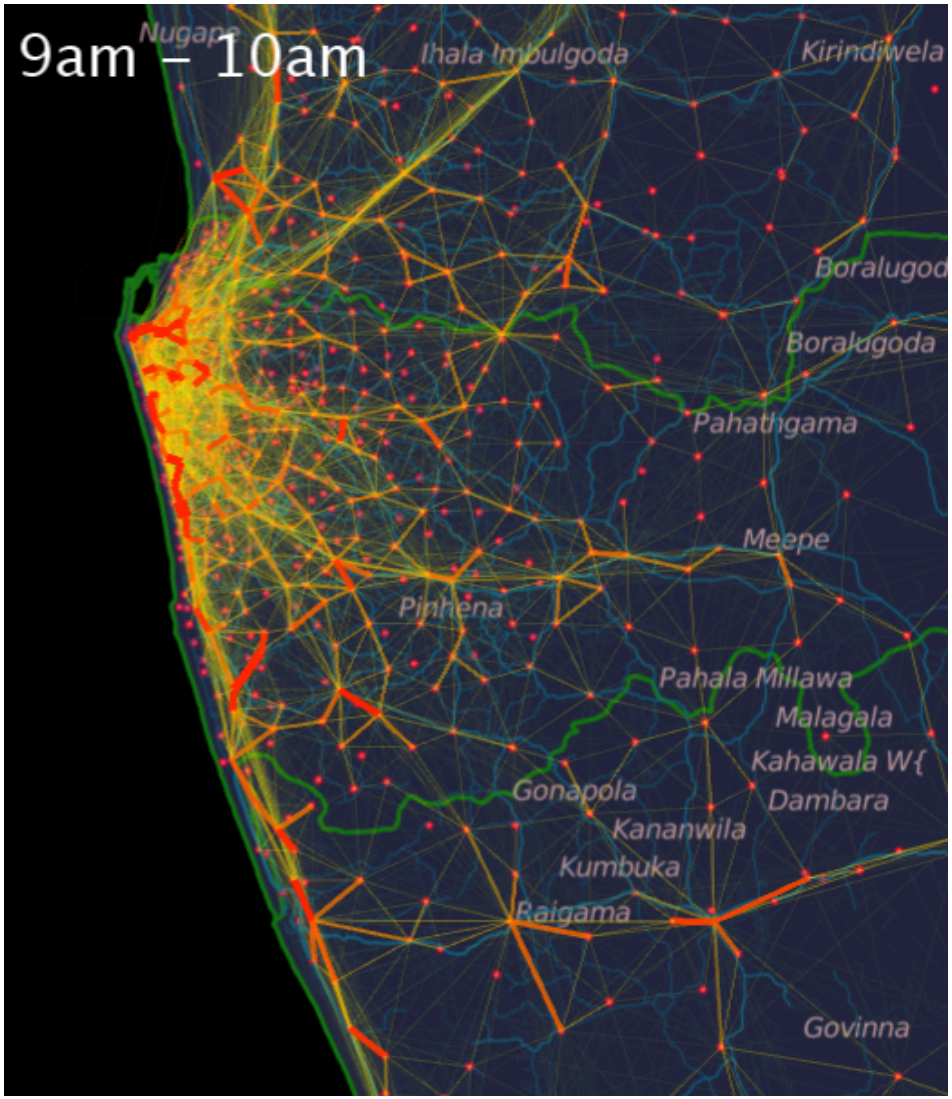
LIRNEasia
Pro-poor. Pro-market.

Population density changes in Colombo region: weekday/ weekend

Pictures depict the change in population density at a particular time relative to midnight



Mobility patterns of the population

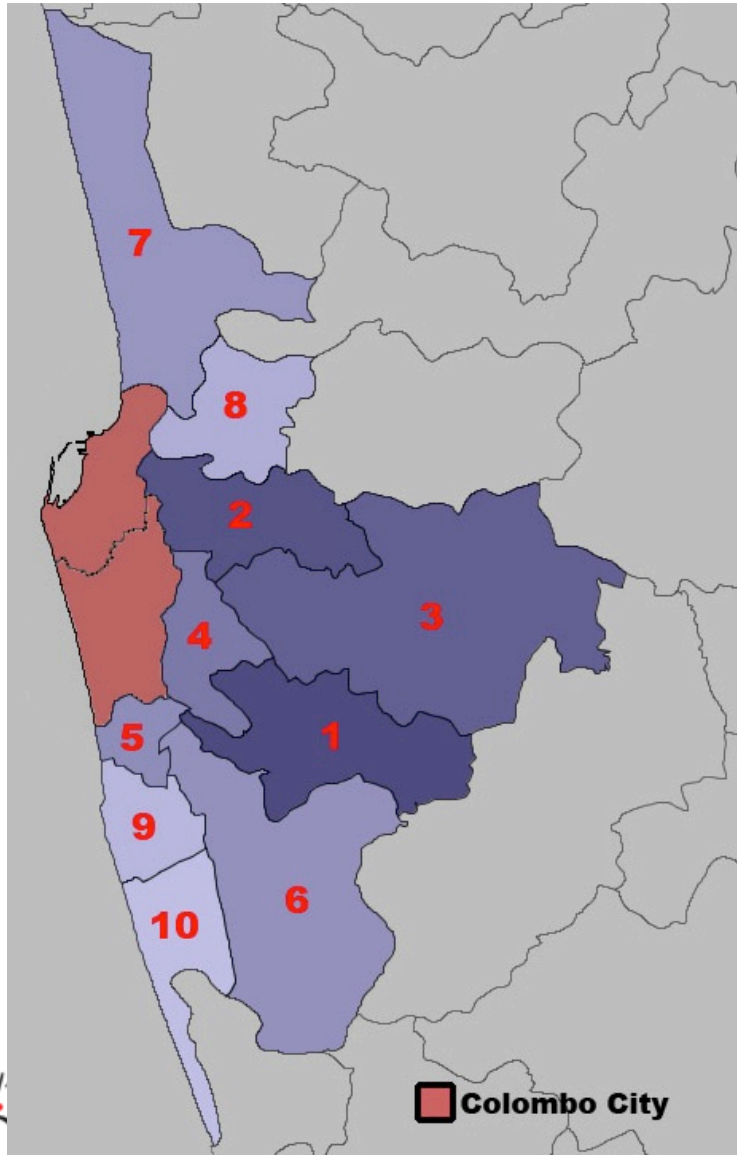


Low → High

Volume of people

LIRNEasia
Pro-poor. Pro-market.

Where people live and work: 46.9% of Colombo City's daytime population comes from the surrounding regions



Colombo city is made up of Colombo and Thimbirigasyaya DSDs

Home DSD	%age of Colombo's daytime population
Colombo city	53.1
1. Maharagama	3.7
2. Kolonnawa	3.5
3. Kaduwela	3.3
4. Sri Jayawardanapura Kotte	2.9
5. Dehiwala	2.6
6. Kesbewa	2.5
7. Wattala	2.5
8. Kelaniya	2.1
9. Ratmalana	2.0
10. Moratuwa	1.8

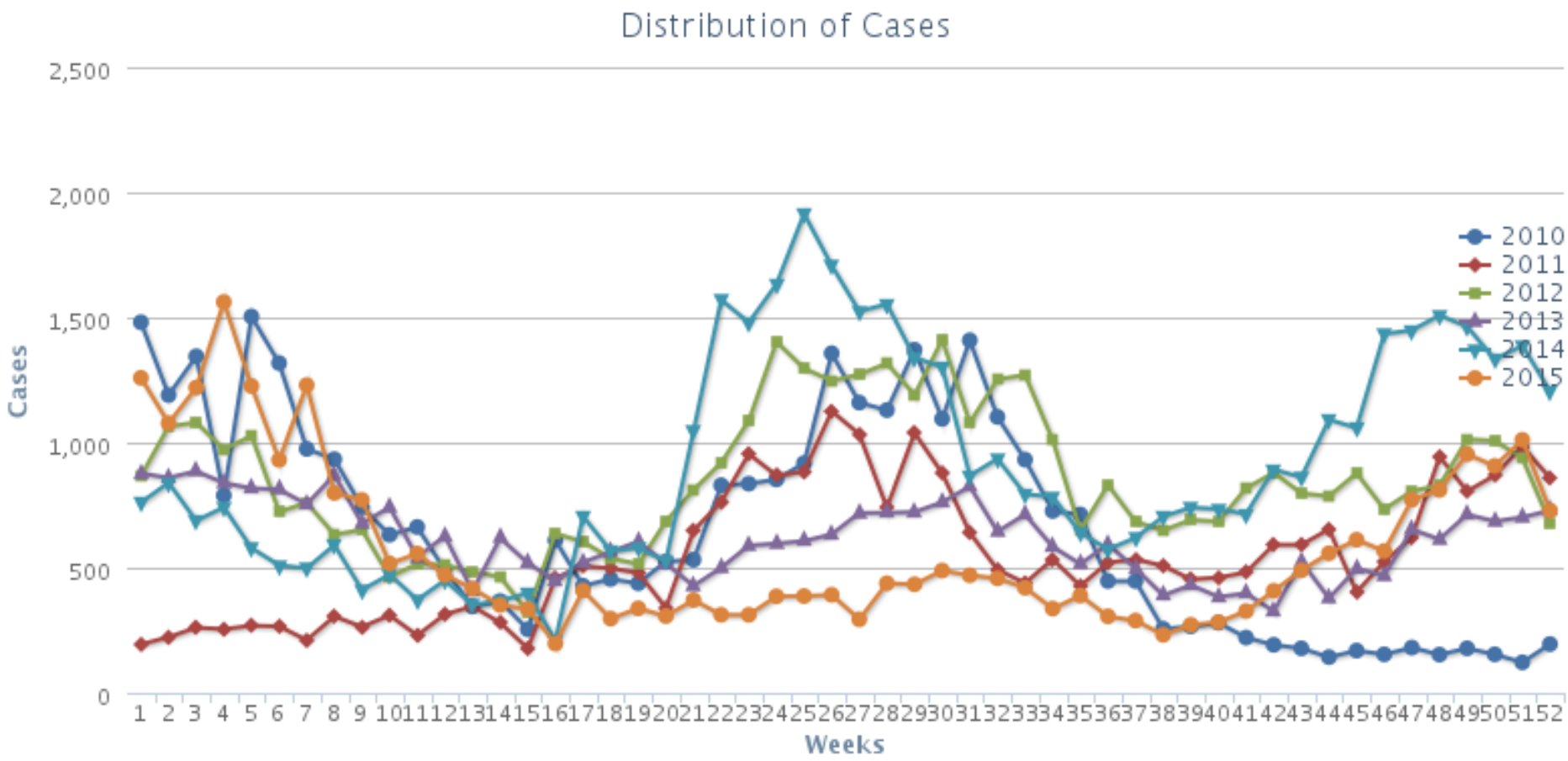
Other ongoing research

- Modeling infectious disease propagation (Dengue) based on human mobility from CDR
- Measuring the impact of a transport shock (Opening the E03 expressway)
- Traffic analysis using CCTV footage
- Enhancing land use predictions with social media data (Eg. Foursquare), satellite imagery, and mobile network big data
- Modeling interactions between different land use patterns
- Developing new high-frequency indicators of economic activity
- Developing socio-economic indices and poverty mapping using CDR, satellite imagery, census data, etc.

Modeling infectious disease propagation (Dengue) based on human mobility from CDR

- Co-supervised a student group from University of Moratuwa with Dr Amal Shehan Perera in 2015/16

Dengue in Sri Lanka: Temporal trend



Source: http://www.epid.gov.lk/web/index.php?option=com_casesanddeaths&Itemid=448&lang=en#

Dengue in 2015: Spatial distribution of notified cases

RDHS	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	**Total**
Colombo	1802	1101	537	317	515	515	907	711	401	769	985	1321	**9881**
Gampaha	817	492	255	193	264	225	317	233	186	277	387	496	**4142**
Kalutara	276	180	106	72	86	78	125	92	57	118	167	202	**1559**
Kandy	253	156	95	63	71	70	64	47	62	154	120	170	**1325**
Matale	166	92	26	16	10	16	8	7	8	12	18	22	**401**
N Eliya	38	26	13	5	8	8	15	5	6	15	14	27	**180**
Galle	150	118	70	35	26	20	46	80	75	131	151	128	**1030**
Hambantota	46	51	28	18	12	11	26	27	17	61	56	45	**398**
Matara	81	46	41	28	25	12	21	24	24	72	43	42	**459**
Jaffna	614	218	113	48	66	63	69	45	51	76	216	437	**2016**
Kilinochchi	18	11	3	0	2	4	9	7	7	14	11	6	**92**
Mannar	50	16	2	1	4	2	1	1	2	2	4	20	**105**
Vavuniya	32	15	8	4	15	9	5	4	11	19	16	59	**197**
Mulativu	36	17	8	10	12	14	13	7	2	2	6	15	**142**
Batticaloa	335	347	210	124	139	82	65	18	8	31	40	75	**1474**
Ampara	10	5	6	0	5	9	4	6	5	3	3	11	**67**
Trincomalee	131	96	59	81	65	39	30	9	8	6	15	48	**587**
Kurunegala	329	178	108	73	62	61	82	70	27	56	107	100	**1253**
Puttalam	256	85	34	20	32	42	65	14	12	23	77	79	**739**
Apura	122	78	27	18	12	15	18	13	14	12	29	43	**401**
Polonnaruwa	47	48	13	9	12	9	18	12	8	21	25	28	**250**
Badulla	195	60	30	31	22	30	33	19	10	35	54	47	**566**
Moneragala	46	29	15	7	7	18	16	14	2	16	23	30	**223**
Ratnapura	163	122	90	59	70	68	105	74	46	64	93	87	**1041**
Kegalle	104	57	36	40	54	38	50	49	36	70	80	97	**711**
Kalmunai	228	87	29	21	29	19	13	16	14	7	22	53	**538**
TOTAL	**6345**	**3731**	**1962**	**1293**	**1625**	**1477**	**2125**	**1604**	**1099**	**2066**	**2762**	**3688**	**29777**

Source: http://www.epid.gov.lk/web/index.php?option=com_casesanddeaths&Itemid=448&lang=en#