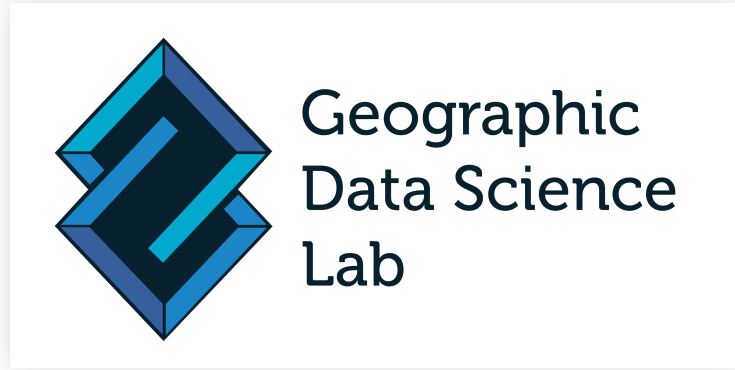
Images as (policy-ready) data

Dani Arribas-Bel [@darribas]



Policy Data & Images

Instituto Nacional de Estadística

iria Colaboradores en encuestas del INE
INEbase La información estadística
+ Agricultura y medio ambiente
+ Ciencia y tecnología
+ Demografía y población
+ Economía
+ Industria, energía y construcción
+ Mercado laboral
+ Nivel y condiciones de vida (IPC)
+ Servicios
+ Sociedad
Síntesis estadística
Estadísticas territoriales
Información estadística europea
Indicadores económicos. FMI
ODS. Indicadores Agenda 2030

iCal **Calendario** 2019

Noviembre

<<						>>
L	M	X	J	V	S	D
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

Calidad y Buenas Prácticas
másINE Revista digital
el Explica
17 OBJETIVOS DE DESARROLLO SOSTENIBLE
El IPC en un clic
Apellidos y nombres
¿Cuántos habitantes...?

Acceso a todos nuestros vídeos e infografías

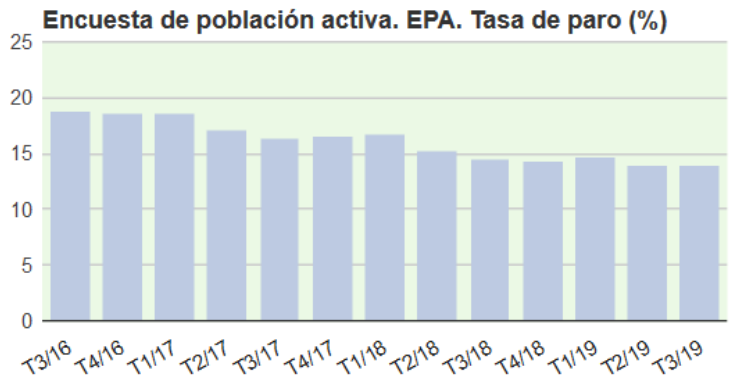
Última hora

22 Nov 19. Coyuntura turística hotelera. CTH
21 Nov 19. Entrada de pedidos en la industria. IEP
21 Nov 19. Indices de cifras de negocios en la Industria. ICN
21 Nov 19. Indicadores de actividad del sector servicios. IASS
20 Nov 19. Demografía armonizada de empresas
19 Nov 19. Estadística del taxi

Más noticias ...

Indicador		Periodo	Valor	Variación (%)
IPC		2019M10	105,126	0,1
EPA. Ocupados (miles)		2019T3	19.874,3	1,77
EPA. Tasa de paro	1	2019T3	13,92	-0,63
PIB	2 4	2019T3	110,9	2,0
Población total (miles)	3	2019	46.934,6	0,59

1. Valor en %. Variación: diferencia respecto a la tasa del mismo período del año anterior
2. Índice volumen encadenado, ref. 2015. Datos corregidos de efectos estacionales y de calendario
3. Cifras de población a 1 de enero de 2019. Datos provisionales
4. Datos avance



OECD Data

› Français

Find, compare and share the latest **OECD data**: charts, maps, tables and related publications …

Search

Browse by topic or country

Search tips
› Catalogue of OECD databases

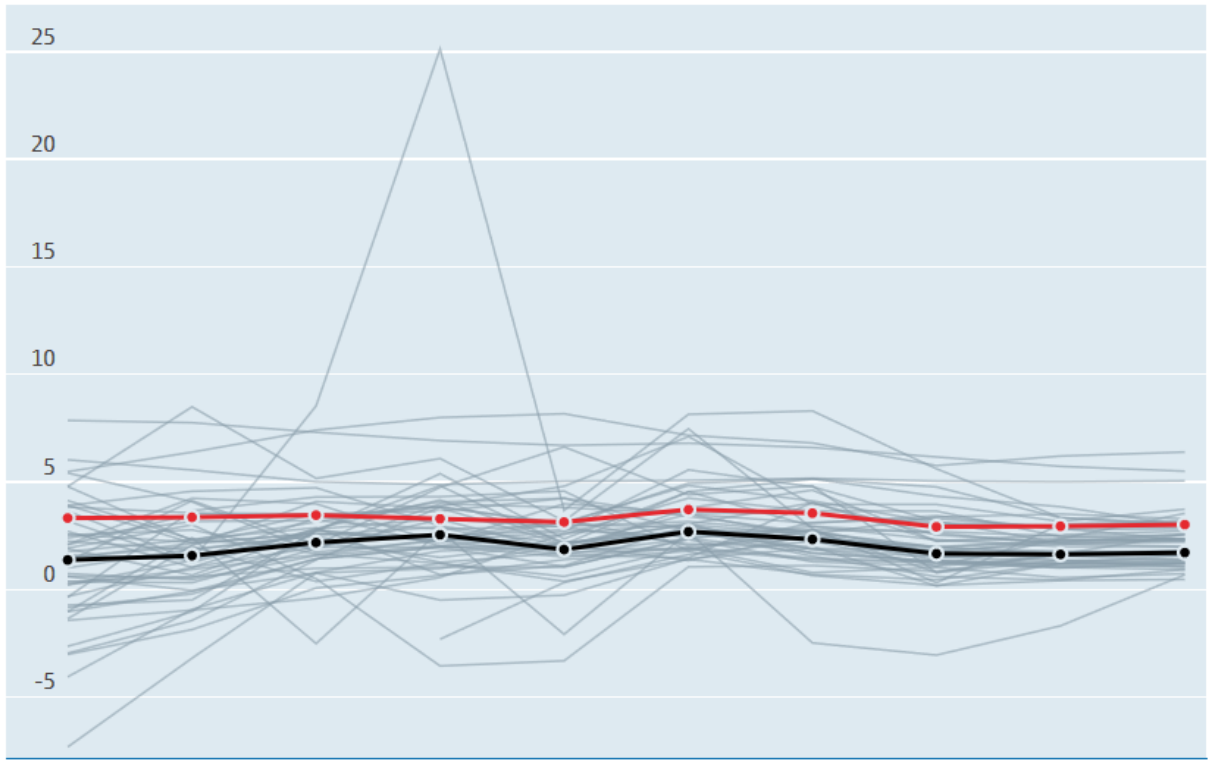
OECD Data

Featured Charts

The global outlook is unstable, see the latest OECD Economic Outlook.

Real GDP forecast

Total, Annual growth rate (%), 2012 – 2021



Latest news

Statistical news releases

See recent statistical news releases.

Compare your income

Try OECD's Compare your income to check your perception of income inequality against reality!

Statistical resources

Database access

Make your own queries on large databases in our data warehouse, OECD.Stat.

37 Calle Gran Vía - Google Ma

https://www.google.com/maps/@40.4202895,-3.7048931,3a,75y,292.72h,94.85t/data=!3m6!1e1!3m4!1s-nhf7

37 Calle Gran Vía
Madrid, Community of Madrid
Google
Street View - Apr 2011

BURGER KING
MENÚ MEDIANO WHOPPER
BURGER KING
BURGER KING
Santander
SERVICIO ESPECIAL A HOSTELERÍA
FRUTAS LOGROÑO
91 - 601 72 28
91 - 601 85 17
TAXI
BUS
Calle Gran Vía
Callao
Google
Image capture: Apr 2011 © 2019 Google United Kingdom Terms Report a problem

Planet Explorer

https://www.planet.com/explorer/#/labels/0/cloud_cover/0,0.05/center/-3.735,40.401/zoom/13/geometry

Madrid

2019/08/23 - 2019/11/23 Save search

Browse Compare Stories

Daily Imagery

Get more from Planet Explorer
Browse the Map and Timeline to find additional imagery available for purchase.

Cloud cover 0 – 5% | Area coverage 0 – 100% | Source 3 sources | All filters >

31 of many Most recent ↓

Oct 26, 2019 11:02:14 UTC
4-band PlanetScope Scene (3 m)
92% area coverage
64 images

Oct 25, 2019 10:59:42 UTC
4-band PlanetScope Scene (3 m)
71% area coverage
40 images

Oct 24, 2019 13:33:10 UTC
4-band PlanetScope Scene (3 m)
23% area coverage
23 images

Oct 21, 2019 10:59:36 UTC
4-band PlanetScope Scene (3 m)
34% area coverage
21 images

Oct 19, 2019 10:59:30 UTC
4-band PlanetScope Scene (3 m)
1% area coverage
1 image

Oct 16, 2019 10:40:06 UTC
4-band PlanetScope Scene (3 m)
31% area coverage
20 images

API {:} ORDER ITEMS (133)

Terms

NDVI

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Weekly Basemap
Daily Imagery

40.43° N, 3.68° W 13 1000 m

Why Images?

- Visual cues encode of socio-economic information
- Largely unexplored (in Social Sciences)
- Nicely complement other available data

Why now?

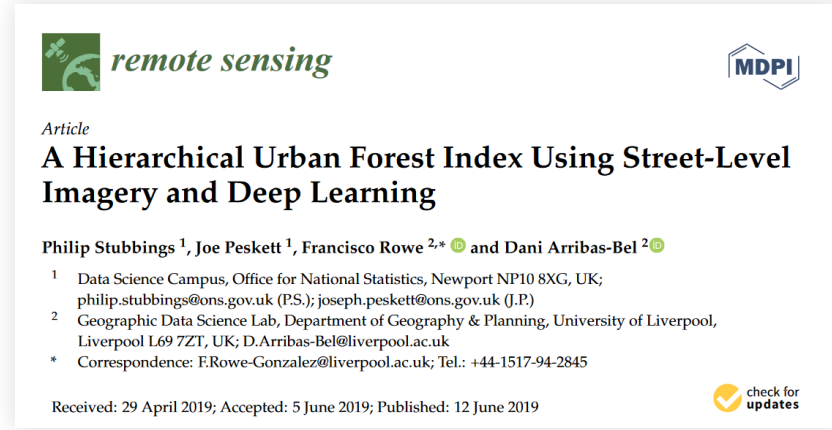
- There's a lot of them
- They get better every day
- Technological opportunism

Some examples…

Measuring Green Coverage

Measuring Green Coverage

- Urban green space has many benefits to urban dwellers
- But is is hard/expensive to measure
- Better quantification can inform evidence-driven policies



remote sensing

MDPI

Article

A Hierarchical Urban Forest Index Using Street-Level Imagery and Deep Learning

Philip Stubbings [1], Joe Peskett [1], Francisco Rowe [2,*] and Dani Arribas-Bel [2]

[1] Data Science Campus, Office for National Statistics, Newport NP10 8XG, UK; philip.stubbings@ons.gov.uk (P.S.); joseph.peskett@ons.gov.uk (J.P.)

[2] Geographic Data Science Lab, Department of Geography & Planning, University of Liverpool, Liverpool L69 7ZT, UK; D.Arribas-Bel@liverpool.ac.uk

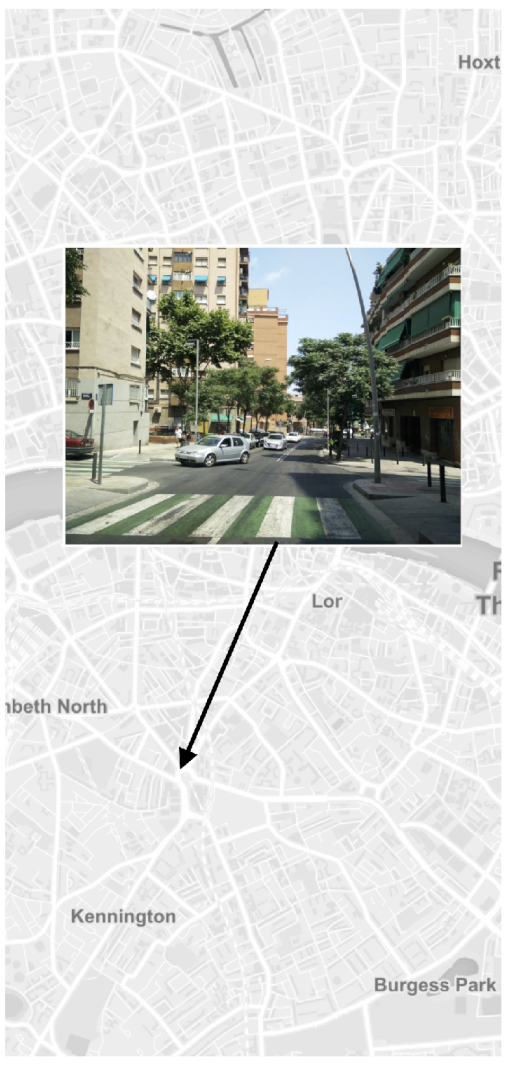
* Correspondence: F.Rowe-Gonzalez@liverpool.ac.uk; Tel.: +44-1517-94-2845

Received: 29 April 2019; Accepted: 5 June 2019; Published: 12 June 2019

check for updates

Input Data

Google Street View (GSV) Imagery



Stage 1

Vegetation Pixel Identification & Classification

Methods

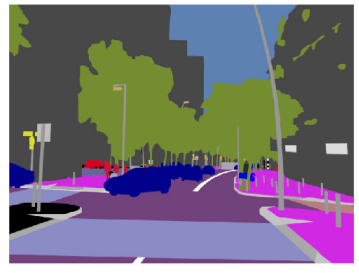
1. Threshold Methods

L^*a^*b

Random Forest

2. Semantic Segmentation

Pyramid Scene Parsing Network (PSPNet)



Pixel Class

Binary

- Green
- Non-green

Multiple

- Tree
- Grass
- Plant
- Earth
- Sky
- Car
- Boat
- Water
- River
- House
- Building
- Skyscraper
- Wall
- Floor

Pixel Classification

Image ID	Vegetation Pixel*
1	1
1	0
1	1
⋮	⋮
1	1
2	0
2	0
2	1
⋮	⋮
2	1
⋮	⋮
n	1
n	1
n	1
⋮	⋮
n	0

*1: Yes; 0: No

Stage 2

Aggregation Score

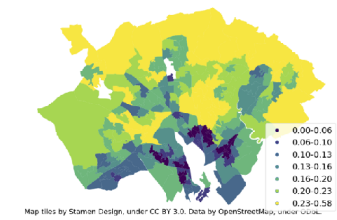
Hierarchical Regression Model

Level 1: Image level

Level 2: Geographic area level

Level 2: Area ID	Level 1: Image ID	Vegetation Pixel*
1	1	1
1	1	0
1	1	1
⋮	⋮	⋮
1	1	1
1	2	0
1	2	0
1	2	1
⋮	⋮	⋮
1	2	1
⋮	⋮	⋮
n	n	1
n	n	1
n	n	1
⋮	⋮	⋮
n	n	0

Hierarchical Tree Score



Outcome

So what?

- Information on exposure to vegetation in urban environments is hard to generate…
- … But very important for a variety of challenges, from pollution to mental health
- (Geographic) Data Science can help produce timely datasets at scale

Aerial/Satellite Imagery

*The **physical appearance** of a human settlement is a **reflection** of the **society** that inhabits it*

*Urban areas with similar **built environment** have similar **social and demographic** characteristics*

Remote sensing-based
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0176684
plos.org create account sign in
PLOS ONE
PUBLISH ABOUT BROWSE
SEARCH
advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Remote sensing-based measurement of Living Environment Deprivation: Improving classical approaches with machine learning

Daniel Arribas-Bel, Jorge E. Patino, Juan C. Duque

Published: May 2, 2017 • https://doi.org/10.1371/journal.pone.0176684

18 Save
3 Citation
3,508 View
5 Share

Article Authors Metrics Comments Media Coverage
Download PDF
Print Share

Abstract
Introduction
Data
Methods
Results
Discussion
Conclusions
Author Contributions
References

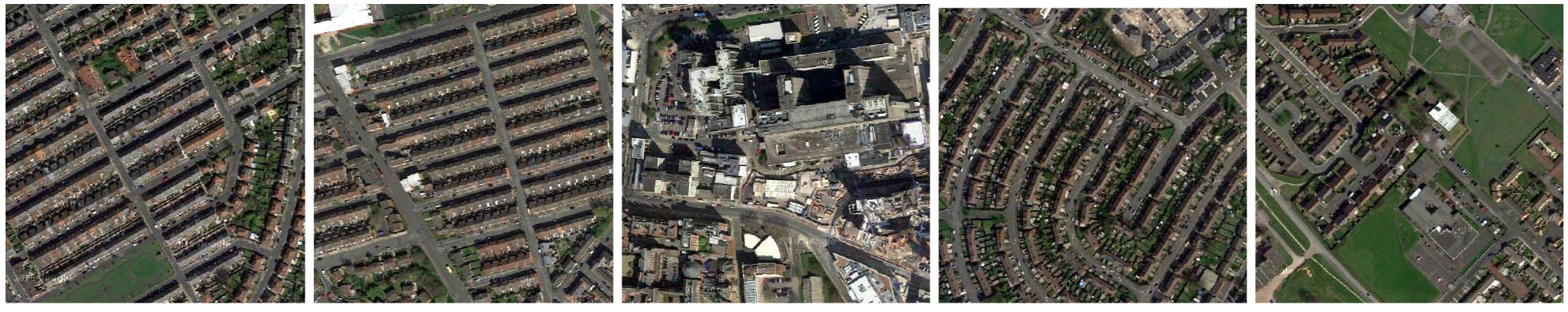
Reader Comments (0)
Media Coverage (0)

Abstract

This paper provides evidence on the usefulness of very high spatial resolution (VHR) imagery in gathering socioeconomic information in urban settlements. We use land cover, spectral, structure and texture features extracted from a Google Earth image of Liverpool (UK) to evaluate their potential to predict Living Environment Deprivation at a small statistical area level. We also contribute to the methodological literature on the estimation of socioeconomic indices with remote-sensing data by introducing elements from modern machine learning. In addition to classical approaches such as Ordinary Least Squares (OLS) regression and a spatial lag model, we explore the potential of the Gradient Boost Regressor and Random Forests to improve predictive performance and accuracy. In addition to novel predicting methods, we also introduce tools for model interpretation and evaluation such as feature importance and partial dependence plots, or cross-validation. Our results show that Random Forest proved to be the best model with an R^2 of around 0.54, followed by Gradient Boost Regressor with 0.5. Both the spatial lag model and the OLS fall behind with significantly lower performances of 0.43 and 0.3, respectively.

Check for updates
ADVERTISEMENT

RSF 1.31 1.35 1.40 1.45 1.51

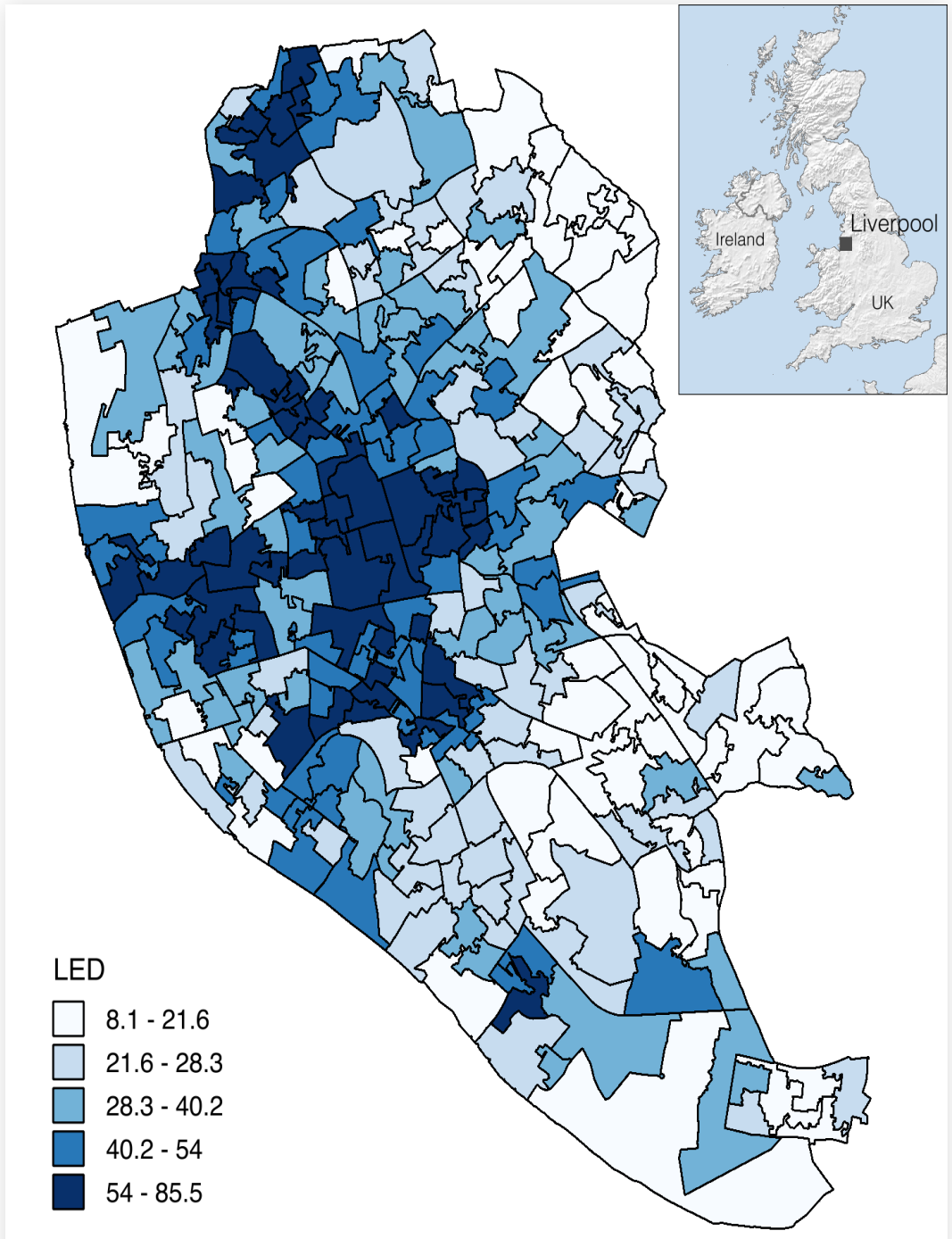


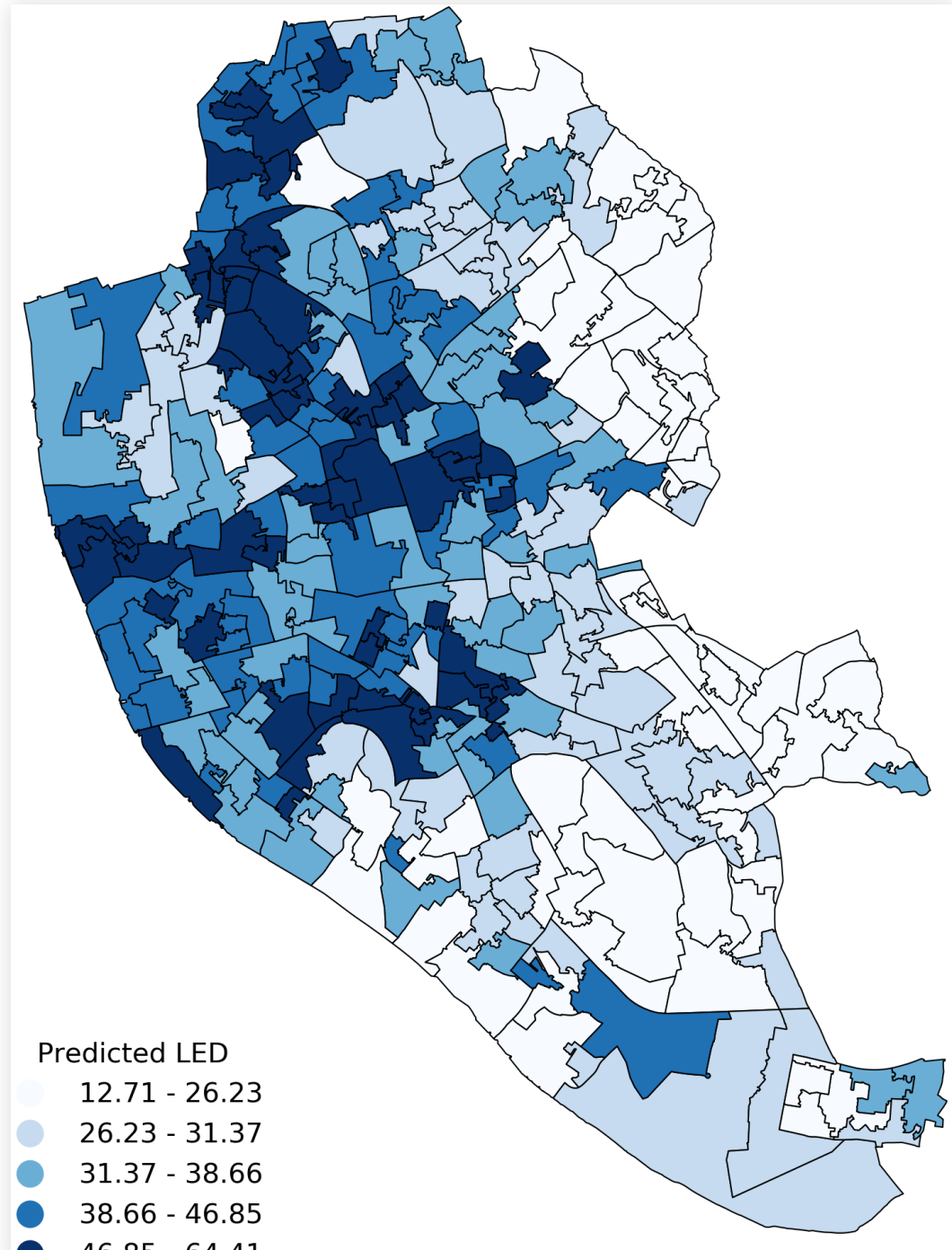
p_veg 5 % 20 % 35 % 50 % 65 %

p_b_water 0.1 % 2 % 4 % 6 % 8 %

RMM 1.16 1.23 1.30 1.37 1.41

0 100 m


Ireland
Liverpool
UK
LED
8.1 - 21.6
21.6 - 28.3
28.3 - 40.2
40.2 - 54
54 - 85.5


Predicted LED
12.71 - 26.23
26.23 - 31.37
31.37 - 38.66
38.66 - 46.85
46.85 - 64.41

So what?

- Generating an update of IMD is expensive
- Satellite images are *already* being collected
- "Intercensal" updates based on satellites?

To take away

- There's a *lot* of images (and more to come!)
- Images *can* be useful for policy analysis
- To make the most of them, you need (Geographic) Data Science

Images as (policy-ready) data

Dani Arribas-Bel [@darribas]

UNIVERSITY OF LIVERPOOL

Geographic Data Science Lab