



Modelagem da ilha urbana de calor em Londrina usando aprendizado de máquina

Modeling the urban heat island in the city of Londrina: A machine learning approach

Gabriel Yoshikazu Cortez Oukawa*, Admir Créso de Lima Targino[†],

Patricia Krecl Abad[‡]

RESUMO

Caracterizar a variabilidade espaço-temporal da intensidade da Ilha Urbana de Calor (IIUC) é fundamental para o desenvolvimento de cidades resilientes a mudanças climáticas. Este estudo simulou a evolução espaço-temporal da IIUC na cidade de Londrina (PR) durante o verão de 2011–2012 usando uma combinação de dados *in situ* e sensoriamento remoto, com modelagem de aprendizado de máquina; mais especificamente, florestas aleatórias (FA). As FA identificaram as principais variáveis preditoras da temperatura do ar (T_{ar}) para os períodos diurno e noturno. O modelo de FA diurno explicou 96% da variância observada e identificou a pressão atmosférica, velocidade do vento e irradiância solar como principais preditores; já o noturno explicou 89% da variância, e identificou a velocidade do vento, pressão atmosférica e cobertura de nuvens como principais preditores. Subsequentemente, as FA foram aplicadas para mapear a IIUC em escala espacial fina (10 m) nos períodos diurno e noturno. Nossos resultados sugerem que modelos baseados em FA podem servir como base para o desenvolvimento de estratégias que visam reduzir a exposição pessoal a altas T_{ar} .

Palavras-chave: Clima urbano; Florestas aleatórias; Sensoriamento remoto; Aprendizado de máquina

ABSTRACT

Characterizing the spatiotemporal variability of the Urban Heat Island intensity (UHII) is a key step in developing cities that are resilient to climate change. In this study, we simulated the spatiotemporal evolution of the UHII in Londrina (Paraná) during the summer of 2011–2012 by using a combination of *in situ* data, remote sensing and machine learning models; more specifically, random forests (RF). RF identified the main predictors of the air temperature (T_{air}) during daytime and nighttime conditions. The daytime model explained 96% of the observed variance and identified atmospheric pressure, wind speed and solar irradiance as its main predictors; the nighttime model explained 89% of the variance and identified wind speed, atmospheric pressure and cloud cover as the main explanatory variables. Subsequently, the RF were applied to map out the UHII at a fine spatial scale (10 m) for both daytime and nighttime periods. Our results suggest that RF models can provide a robust framework for developing city-level strategies aimed at reducing personal exposure to scorching T_{air} .

Keywords: Urban climate; Random forests; Remote sensing; Machine learning

* Engenharia Ambiental e Sanitária, Universidade Tecnológica Federal do Paraná, Londrina, Paraná, Brasil; gabrieloukawa@alunos.utfpr.edu.br

[†] Universidade Tecnológica Federal do Paraná, Campus Londrina; admirtargino@utfpr.edu.br

[‡] Universidade Tecnológica Federal do Paraná, Campus Apucarana; patriciak@utfpr.edu.br



1 INTRODUÇÃO

A crescente expansão urbana, e troca de superfícies naturais por artificiais provoca alterações nas propriedades aerodinâmicas e radiativas das superfícies (rugosidade, albedo e emissividade), propriedades térmicas do substrato (inércia, condutividade e admitância térmica), propriedades hidráulicas do solo (permeabilidade e capacidade hidráulica) e introduzem grandes quantidades de calor antropogênico (OKE *et al.*, 2017). Como consequência, o balanço energético da área urbana difere substancialmente daquele em áreas rurais; isso provoca alterações no clima local e intensifica a ilha urbana de calor (IUC). Esse fenômeno ocorre quando a temperatura da atmosfera urbana é maior do que das áreas rurais ao seu redor (OKE *et al.*, 2017).

Embora a inserção e/ou preservação de áreas verdes seja um caminho eficiente para mitigar altas temperaturas em ambientes urbanos, esse argumento pode ser muito genérico, e não leva em conta a influência de parâmetros locais no desenvolvimento da IUC (MANOLI *et al.*, 2019). Nesse sentido, modelos estatísticos têm sido amplamente utilizados para entender e simular o efeito de materiais artificiais (e seus arranjos e disposições geométricas) no comportamento da IUC em diferentes escalas espaço-temporais. Esses modelos normalmente incluem um grande número de variáveis independentes (VI, espaciais e temporais), que são usadas para simular o comportamento de uma variável dependente (VD, também conhecida como “variável resposta”). As florestas aleatórias (FA) é uma dessas técnicas que, embora utilizada em diversas áreas de conhecimento ao redor do mundo (CUTLER *et al.*, 2007; BECHTEL; DANEKE, 2012), ainda é pouco explorada em estudos sobre a IUC.

O presente estudo analisou a evolução espaço-temporal da IUC na cidade de Londrina (PR) durante o verão de 2011–2012 usando uma combinação de dados *in situ*, sensoriamento remoto e FA. Mais especificamente, inclui dados de temperatura do ar coletados por uma rede urbana, variáveis meteorológicas (velocidade do vento, precipitação, cobertura de nuvens, etc.), uso e cobertura do solo, e índices urbanos (densidade populacional, altura dos edifícios, etc.). Os modelos de FA foram aplicados para o período diurno e noturno usando a temperatura do ar como variável resposta, e simularam a intensidade da Ilha Urbana de Calor (IIUC) em fina escala espacial (10 m).

2 METODOLOGIA

2.1 Instrumentação e coleta de dados

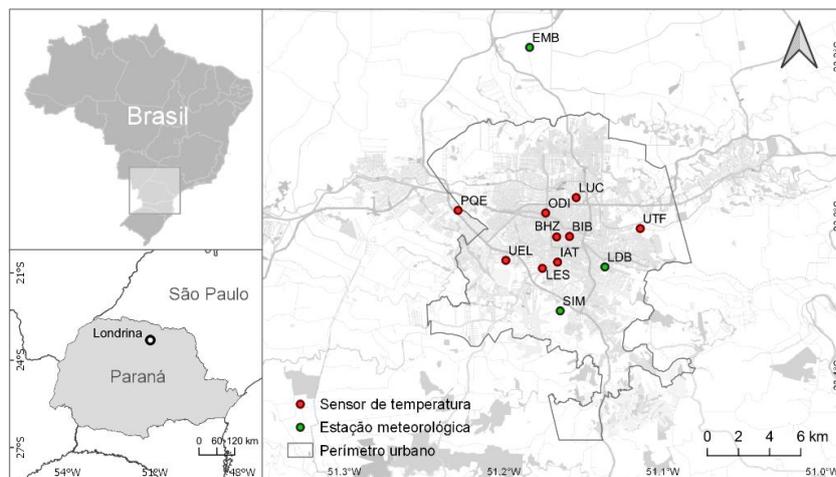
O projeto ICALON (acrônimo de Ilha de Calor de Londrina) foi conduzido de 24 de dezembro de 2011 a 26 de fevereiro de 2012 com o objetivo de observar e mapear o desenvolvimento da IUC em Londrina. A temperatura do ar (T_{ar}) foi monitorada em 12 locais da cidade, nove dos quais (Figura 1, pontos vermelhos) utilizando sensores de temperatura HOBO U23-001 (Onset Computer Corporation, EUA). Nos outros três locais (Figura 1, pontos verdes), a temperatura foi monitorada por estações meteorológicas permanentes.

Figura 1 – Locais de monitoramento *in situ*



SEI-SICITE 2021

Pesquisa e Extensão para um mundo em transformação



Fonte: A autoria própria (2021).

Outros dados meteorológicos horários disponibilizados pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) e pelo Sistema de Tecnologia e Monitoramento Ambiental do Paraná (SIMEPAR) também foram usados, sendo eles: velocidade do vento (VV), direção do vento (DV), pressão atmosférica (P), irradiância solar (RAD) e precipitação acumulada (PA). Adicionalmente, dados horários de cobertura de nuvens (CN) medidos no aeroporto de Londrina (LDB) foram obtidos por mensagens em código METAR (<http://bancodedados.cptec.inpe.br/>).

O uso e cobertura do solo foi caracterizado a partir de imagens de alta resolução do satélite Sentinel-2 (<https://scihub.copernicus.eu/>). Modelos de elevação digital da *Shuttle Radar Topography Mission* (<https://earthexplorer.usgs.gov/>) foram usados para determinar a elevação em cada local. Adicionalmente, a altura dos edifícios e comprimento das vias foram providos pelo Instituto de Pesquisa e Planejamento Urbano de Londrina (IPPUL), a distribuição espacial dos pontos de ônibus foi fornecida pela Companhia Municipal de Trânsito e Urbanização de Londrina (CMTU) e dados populacionais pelo Instituto Brasileiro de Geografia e Estatística (IBGE, <https://mapas.ibge.gov.br/interativos/grade/>).

2.2 Análise, tratamento e processamento de dados

As variáveis espaciais foram agrupadas em cinco categorias: uso e cobertura do solo, topografia, morfologia urbana, e população e tráfego de veículos (Tabela 1). Foram criados *buffers* (polígonos circulares com raio específico em torno de uma feição) com tamanhos de 100, 200, 300, 400, 500 e 1.000 m para cada uma das 12 estações de monitoramento e os dados espaciais foram agregados a cada respectivo *buffer*. Usando classificação supervisionada, o uso e cobertura do solo foi dividido em cinco classes principais dentro da área de estudo e agregado aos *buffers*, com as seguintes frações: asfalto (ASFT), telhado (TLHD), vegetação (VEG), água (AGUA) e solo exposto (SOLO). Adicionalmente, o fator de visão do céu (FVC), altura média (H) e densidade volumétrica dos edifícios (DVE) foram calculados e agregados em cada *buffer*. O comprimento das vias (CVIA) e número de pontos de ônibus (BUS) contidos em cada *buffer* foram utilizados como uma estimativa do tráfego de total de veículos e de veículos pesados, respectivamente. A densidade populacional (POP) também foi agregada aos *buffers*, que junto ao comprimento das vias e número de pontos de ônibus podem servir como indicador da quantidade de calor antropogênico produzido na área de estudo por fontes móveis e metabolismo humano (HANKEY; MARSHALL, 2015).



Para os cálculos da IIUC, a estação UEL foi utilizada como referência (é a que melhor captura os padrões de vegetação nativa da região) e a estação BHZ como representante da área urbana (que no geral demonstrou ser a mais quente) (TARGINO *et al.*, 2014).

2.3 Modelo de florestas aleatórias

O desenvolvimento do modelo de FA consiste no uso de um conjunto de VI para modelar/simular uma variável resposta (VD). O modelo de FA é uma técnica de aprendizado de máquina não paramétrica e não linear (BREIMAN, 2001), construído a partir da média de preditores treinados usando árvores de classificação ou regressão. Por se tratar de uma técnica não linear, não há nenhum pré-requisito matemático ou estatístico para as VI e VD (*e.g.*, normalidade, linearidade, homocedasticidade).

Dois modelos foram desenvolvidos utilizando a T_{ar} como variável resposta; um diurno e outro noturno. A variável RAD foi usada como parâmetro para filtrar o conjunto de dados nessas duas categorias, sendo que o modelo diurno utiliza dados coincidentes com $RAD > 0$ e o noturno utiliza dados coincidentes com $RAD = 0$. A divisão entre os dois períodos é baseada em estudos realizados em Londrina que apontaram diferenças significativas entre a IUC diurna e noturna (TARGINO *et al.*, 2014; ANJOS *et al.*, 2020). O desempenho de cada modelo foi avaliado através do coeficiente de determinação (R^2) e raiz do erro quadrático médio (RMSE, em inglês) obtidos por validação cruzada *k-fold*.

Tabela 1 – Conjunto de VI usadas na construção dos modelos

Categoria	Variável	Abreviação	Unidade
Uso e cobertura do solo	Fração de asfalto	ASFT	%
	Fração de telhado	TLHD	%
	Fração de vegetação	VEG	%
	Fração de água	AGUA	%
	Fração de solo exposto	SOLO	%
Topografia	Elevação	Z	m
Geometria urbana	Fator de visão do céu	FVC	%
	Altura média dos edifícios	H	m
	Densidade volumétrica dos edifícios	DVE	%
População e tráfego	Densidade populacional	POP	hab. km ⁻²
	Comprimento das vias	CVIA	m
	Número de pontos de ônibus	BUS	número
Meteorologia	Umidade relativa	UR	%
	Velocidade do vento	VV	m s ⁻¹
	Direção do vento	DV	setores
	Irradiância solar	RAD	W m ⁻²
	Pressão atmosférica	P	hPa
	Precipitação acumulada	PA	mm
	Cobertura de nuvens	CN	%

Fonte: Autoria própria (2021).

3 RESULTADOS

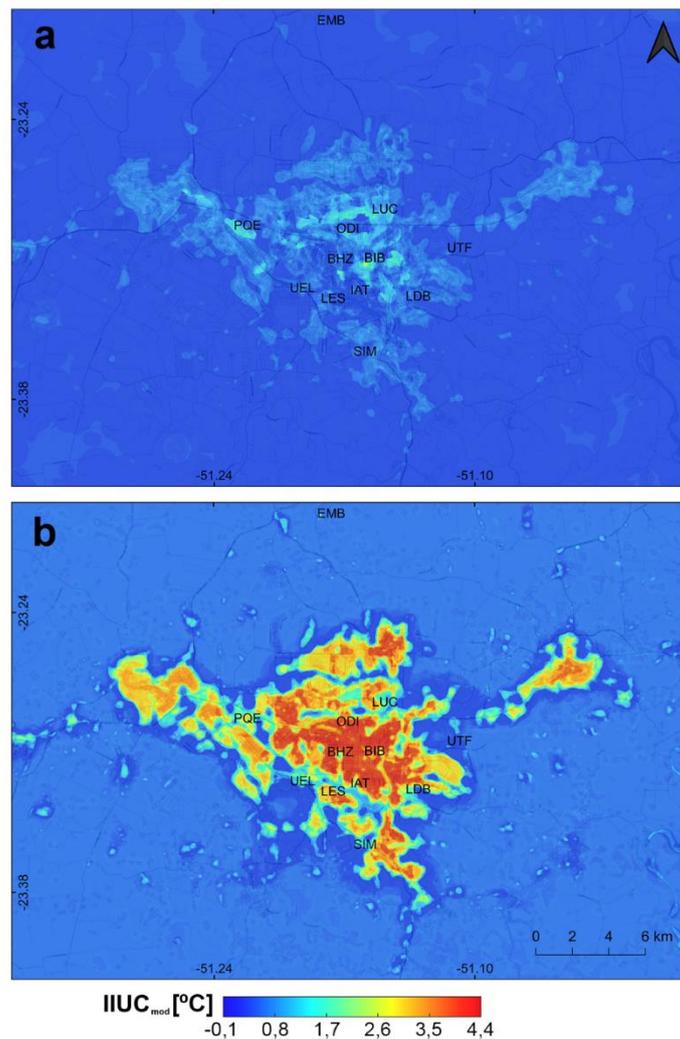
3.1 Resultados dos modelos e IIUC simulada

A validação cruzada do modelo diurno resultou em um R^2 de 0,96 e RMSE de 0,77 °C. As cinco variáveis mais importantes para a predição da T_{ar} (obtidas pela permutação das observações *out-of-bag*) foram: P, VV, RAD, UR e ASFT (no *buffer* de 1.000 m). O modelo noturno resultou em um R^2 de 0,89 e RMSE de 0,90 °C,

com as seguintes variáveis mais importantes: VV, P, CN, UR e DV. Considerando o bom desempenho de ambos os modelos (*i.e.*, explicaram a maior parte da variância da T_{ar} com erros relativamente baixos), eles foram aplicados para simular a evolução espaço-temporal da IIUC em fina escala (10 m e horária).

Dois mapas foram gerados para o dia em que foi observada a maior IIUC (07 de fevereiro, Fig. 3); um para representar o período diurno e outro para o noturno (12 e 23h, respectivamente).

Figura 3 – Simulação da IIUC de Londrina para o dia 07 de fevereiro de 2012 nos períodos diurno (a, 12h) e noturno (b, 23h), ambos painéis compartilham a mesma escala de IIUC



Fonte: Autoria própria (2021).

A IIUC simulada ($IIUC_{mod}$) variou entre $-0,1$ e $4,4$ °C sobre o perímetro da cidade de Londrina, com intensidades baixas em maior parte da área urbana ($-0,1$ a $1,0$ °C), no período diurno, chegando próximo dos $2,0$ °C em alguns locais do centro (BHZ, BIB) (Fig. 3a). Para o período noturno (Fig. 3b), a $IIUC_{mod}$ teve maior amplitude e valores maiores que $4,0$ °C em toda área central (locais BHZ, BIB e ODI). Nos bairros mais distantes do centro (próximo aos locais PQE, UTF e SIM), a $IIUC_{mod}$ foi ligeiramente menor, se mantendo entre $2,0$ a $3,5$ °C. Intensidades baixas ($< 1,0$ °C) se destacam nos locais IAT e PQE, que estão em parques



com fragmentos de vegetação nativa, indicando que sua presença na área urbana pode contribuir com o desenvolvimento de zonas de ar mais frio ao seu redor.

4 CONCLUSÃO

Os modelos de FA tiveram um bom desempenho e conseguiram explicar 96 e 89% da variância da T_{ar} durante o período diurno e noturno, respectivamente. Além disso, os modelos capturaram a variabilidade espaço-temporal da IIUC em fina escala para toda área de estudo. Esses modelos demonstraram ser uma ferramenta robusta para estimar os efeitos de altas IIUC, além do benefício de não ter nenhum pré-requisito para as variáveis envolvidas no cálculo.

Como os dados utilizados nesse estudo cobrem apenas um conjunto limitado de situações sinóticas, a aplicação dos modelos para outras estações do ano pode levar a interpretações equivocadas. Além disso, mais VI podem ser consideradas na construção dos modelos, como índices de estabilidade atmosférica, radiação de onda longa emitida pelas superfícies e calor emitido por fontes estacionárias (*e.g.*, aparelhos de ar-condicionado e processos industriais).

Do ponto de vista de planejamento urbano, os resultados desses modelos podem servir como base para o desenvolvimento de políticas públicas, visando reduzir a exposição a altas temperaturas por meio de intervenções urbanísticas.

AGRADECIMENTOS

Ao CNPq pela cessão da bolsa de iniciação científica e a Fundação Araucária pelo financiamento dos sensores de temperatura (nº de concessão 470/2010).

REFERÊNCIAS

- ANJOS, M.; TARGINO, A. C.; KRECL, P.; OUKAWA, G. Y.; BRAGA, R. F. Analysis of the urban heat island under different synoptic patterns using local climate zones. **Building and Environment**, 2020.
- BECHTEL, B.; DANEKE, C. Classification of Local Climate Zones Based on Multiple Earth Observation Data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 5, n. 4, p. 1191–1202, 2012.
- BREIMAN, L. Random forests. **Machine Learning**, 2001.
- CUTLER, D. R.; EDWARDS, T. C.; BEARD, K. H.; et al. RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. **Ecology**, v. 88, n. 11, p. 2783–2792, 2007.
- HANKEY, S.; MARSHALL, J. D. Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM2.5, Particle Size) Using Mobile Monitoring. **Environmental Science and Technology**, 2015.
- MANOLI, G.; FATICHI, S.; SCHLÄPFER, M.; et al. Magnitude of urban heat islands largely explained by climate and population. **Nature**, 2019.
- OKE, T. R.; MILLS, G.; CHRISTEN, A.; VOOGT, J. A. **Urban Climates**. Cambridge: Cambridge University Press, 2017.
- TARGINO, A. C.; KRECL, P.; CORAIOLA, G. C. Effects of the large-scale atmospheric circulation on the onset and strength of urban heat islands: a case study. **Theoretical and Applied Climatology**, v. 117, n. 1–2, p. 73–87, 2014.