Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

We used data from three sources. First, we obtained birth records from the Colorado Birth Registry, provided by the Colorado Department of Public Health and Environment under a data use agreement. Second, we downloaded fine particulate matter (PM2.5) air pollution exposure data from the US Environmental Protection Agency (EPA). Third, we downloaded temperature data from the US EPA. The use of these data for the analyses in this paper was approved by the institutional review board of Colorado State University. Data sources are described in Section 2 of the main text.

Availability

☒ Data **are** publicly available

☐ Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available dat*a section, below.

Publicly available data

☒ Data are available online at: https://www.epa.gov/hesc/rsig-related-downloadable-data-files, https://aqs.epa.gov/aqsweb/airdata/download_files.html

☐ Data are available as part of the paper's supplementary material.

☒ Data are publicly available by request, following the process described here:

Birth records data are available upon request from the Colorado Department of Public Health and Environment (CDPHE). In order to obtain the data, a data use agreement (DUA) must be

established with CDPHE and proper IRB approval must be obtained. A data request can be made online at https://www.datarequest.dphe.state.co.us. Information on vital statistics data from CDPHE is available at https://cdphe.colorado.gov/center-for-health-and-environmental-data/registries-and-vital-statistics/vital-statistics-program.

☐ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Discussion of lack of publicly available data:

Description

File format(s)

☒ CSV or other plain text:
☐ Software-specific binary format (.Rda, Python pickle, etc.):
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (described here):

Data dictionary

☐ Provided by the authors in the following file(s):
☐ Data file(s) is (are) self-describiing (e.g., netCDF files)
☒ Available at the following URL:

https://www.epa.gov/hesc/rsig-related-downloadable-data-files
https://aqs.epa.gov/aqsweb/airdata/download_files.html
And by request from the Colorado Department of Public Health and Environment

Additional information (optional)

Part 2: Code

Abstract

The code provided includes scripts to reproduce the data analysis and simulation studies. It also includes scripts to create the data files used in the analyses. A .tar.gz file for the most recent version of the R package to implement the methods is provided in supplemental material. The R package is hosted at https://github.com/danielmork/dlmtree and the reproducibility scripts are hosted at https://github.com/danielmork/HDLM.

The reproducibility repository contains all code in the code directory and has subdirectories for the data processing, data analysis, and simulation. The README file in the root directory explains how to use the scripts.

Description

Code format(s)

☐ Script files
- ☒ R ☐ Python ☐ Matlab
- ☐ Other:

☐ Package
- ☒ R ☐ Python ☐ MATLAB toolbox
- ☐ Other:

☐ Reproducible report
- ☐ R Markdown ☐ Jupyter notebook
- ☐ Other:

☐ Shell script

☐ Other (described here):

Supporting software requirements

Version of primary software used

R 4.3.0

Libraries and dependencies used by the code

1. dlmtree (version 0.9.0)
   * Install from source `install.packages(path_to_file, repos = NULL, type="source")`
   ! Updated software found at github.com/danielmork/dlmtree
2. tidyverse (2.0.0)
3. ggplot2 (3.4.2)
4. viridis (0.6.3)
5. lubridate (1.9.2)
6. data.table (1.14.8)
7. units (0.8.2)
8. sf (1.0.13)
9. tigris (2.0.3)
10. raster (3.6.20)
11. gstat (2.1.1)
12. dplyr (1.1.2)
13. R.utils (2.12.2)

Supporting system/hardware requirements (optional)

none

Parallelization used

☒ No parallel code used
☐ Multi-core parallelization on a single machine/node
Number of cores used:
☐ Multi-machine/multi-node parallelization
Number of nodes and cores used:

License

☐ MIT License (default)
☐ BSD
☒ GPL v3.0
☐ Creative Commons
☐ Other (described here):

Additional information (optional)

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:

☐ Any numbers provided in text in the paper

☒ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))

☐ All tables and figures in the paper

☐ Selected tables and figures in the paper, as explained and justified here:

Workflow details

Location

The workflow is available:

☐ As part of the paper's supplementary material

☒ In this Git repository: https://github.com/danielmork/HDLM

☒ Other:

We have included a copy of a git repo and the R package for the HDLM method as part of the submission. Reproducibility files are available at https://github.com/danielmork/HDLM. The R package is hosted at https://github.com/danielmork/dlmtree.

Format(s)

☐ Single master code file

☐ Wrapper (shell) script(s)

☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

☐ Text file (e.g., a readme-style file) that documents workflow

☐ Makefile

☒ Other (more detail in 'Instructions' below)

Instructions

A mix of R scripts, .Rmd documentation, .txt output, and the manuscript source code from latex.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

☐ <1 minute
☐ 1-10 minutes
☐ 10-60 minutes
☐ 1-8 hours
☒ >8 hours
☐ Not feasible to run on a desktop machine, as described here:

Additional documentation (optional)

Notes (optional)