

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

The paper studies two publicly available datasets to demonstrate the statistical and computational efficiency of the proposed online method. The first dataset is the hourly power consumption data in megawatts from the U.S. Energy Information Administration, covering 8 control zones from June 1, 2013 to August 2, 2018. The second dataset is the traffic accident data from the United States between January 1, 2018 and March 31, 2023, which includes approximately 6.6 million accident records. The two datasets are publicly available at <https://www.eia.gov/electricity> and https://smoosavi.org/datasets/us_accidents.

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 data* section, below.

Publicly available data

- ☒ Data are available online at: <https://www.eia.gov/electricity>; [https://smoosavi.org/datasets/us accidents](https://smoosavi.org/datasets/us_accidents)
☐ Data are available as part of the paper's supplementary material.
☐ Data are publicly available by request, following the process described here:

☐ 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-describing (e.g., netCDF files)
- ☒ Available at the following URL:

<https://www.eia.gov/electricity>; [https://smoosavi.org/datasets/us accidents](https://smoosavi.org/datasets/us%20accidents)

Additional information (optional)

Part 2: Code

Abstract

The provided R code implements the proposed online change rate learning algorithm, which estimates the first-order derivatives of the mean and covariance functions for functional data. Additionally, it includes the implementations of several representative alternatives: the offline kernel method, the offline spline and wavelet methods, and a stochastic gradient descent algorithm. The code is organized in a modular structure and depends on widely used R packages, ensuring the reproducibility of the simulation studies under both dense and non-dense sampling schemes.

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 version 4.3.1 (2023-06-16)
Platform: x86_64-conda-linux-gnu (64-bit)
Running under: CentOS Linux 7 (Core)

Libraries and dependencies used by the code

Matrix products: default

BLAS/LAPACK: /public/home/wangwenwu/anaconda3/envs/r-kdr/lib/libopenblas-r0.3.29.so;

LAPACK version 3.12.0

locale:

[1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8 LC_COLLATE=en_US.UTF-8
[5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=en_US.UTF-8 LC_NAME=C
[9] LC_ADDRESS=C LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

time zone: Asia/Shanghai

tzcode source: system (glibc)

attached base packages:

[1] parallel splines stats graphics grDevices utils datasets
[8] methods base

other attached packages:

[1] dplyr_1.1.3 pracma_2.4.4 wavethresh_4.7.2 doParallel_1.0.17
[5] iterators_1.0.14 foreach_1.5.2 Matrix_1.6-1.1 KernSmooth_2.23-22
[9] fda.usc_2.1.0 mgcv_1.9-0 nlme_3.1-163 fda_6.3.0
[13] deSolve_1.38 fds_1.8 RCurl_1.98-1.12 rainbow_3.8
[17] pcaPP_2.0-3 MASS_7.3-60

loaded via a namespace (and not attached):

[1] compiler_4.3.1 tidyselect_1.2.1 ks_1.14.1 bitops_1.0-7
[5] cluster_2.1.4 lattice_0.22-5 R6_2.6.1 kSamples_1.2-10
[9] generics_0.1.4 tibble_3.2.1 pillar_1.11.0 SuppDists_1.1-9.7
[13] rlang_1.1.1 hrcde_3.4 cli_3.6.1 magrittr_2.0.3
[17] grid_4.3.1 mvtnorm_1.2-3 mclust_6.0.0 lifecycle_1.0.4
[21] vctr_0.6.4 glue_1.6.2 codetools_0.2-20 colorspace_2.1-0
[25] pkgconfig_2.0.3

Supporting system/hardware requirements (optional)

Hygon C86 7375 32-core Processor (2.0 GHz)

Parallelization used

- ☐ No parallel code used
- ☒ Multi-core parallelization on a single machine/node
Number of cores used: 25
- ☐ 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

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

All numerical results reported in this manuscript, including the simulation studies and real data applications, can be fully reproduced using the R code.

The code was developed under R version 4.5.1 and depends on widely used CRAN packages, e.g., `fda`, `fda.usc`, `Kernsmooth`, `mgcv`. All required packages can be installed directly from CRAN.

The simulation studies can be reproduced by running the scripts in the directory `Simulation/`. This directory contains subfolders corresponding to the proposed online change rate learning method and the comparative approaches, including the classical offline kernel method, spline-based method, wavelet-based method, and stochastic gradient descent (SGD) method. Within each subdirectory, scripts labeled with `*_dense_*.R` and `*_sparse_*.R` reproduce the results under the dense and sparse sampling schemes, respectively. Executing these scripts generates the numerical summaries and figures reported in the simulation section of the manuscript.

The real data analyses can be reproduced using the scripts in the directory Realdata/. The two datasets analyzed in the paper are publicly available: the hourly power consumption data from the U.S. Energy Information Administration (<https://www.eia.gov/electricity>) and the U.S. traffic accident dataset (https://smoosavi.org/datasets/us_accidents). After downloading the raw data from these sources, users can preprocess the data and save the corresponding .RData files in the appropriate directory. Executing the analysis scripts will reproduce the empirical results presented in the paper.

Running the above scripts produces the estimated first-order derivatives of the mean and covariance functions, the performance metrics for method comparison, and all figures reported in the manuscript.

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)