Scopus

Document details

Export Download Print E-mail Save to PDF Add to List More...

View at Publisher

Advances in Intelligent Systems and Computing
Volume 884, 2019, Pages 372-386
6th Conference on Information Technologies and Communication of Ecuador, TIC-EC 2018; Riobamba; Ecuador; 21 November 2018 through 23 November 2018; Code 219929

EDA and a Tailored Data Imputation Algorithm for Daily Ozone Concentrations (Conference Paper)

Gualán, R.[a,b], Saquicela, V.[a], Tran-Thanh, L.[b]

[a]Department of Computer Science, University of Cuenca, Cuenca, Ecuador
[b]School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, United Kingdom

Abstract

View references (26)

Air pollution is a critical environmental problem with detrimental effects on human health that is affecting all regions in the world, especially to low-income cities, where critical levels have been reached. Air pollution has a direct role in public health, climate change, and worldwide economy. Effective actions to mitigate air pollution, e.g. research and decision making, require of the availability of high resolution observations. This has motivated the emergence of new low-cost sensor technologies, which have the potential to provide high resolution data thanks to their accessible prices. However, since low-cost sensors are built with relatively low-cost materials, they tend to be unreliable. That is, measurements from low-cost sensors are prone to errors, gaps, bias and noise. All these problems need to be solved before the data can be used to support research or decision making. In this paper, we address the problem of data imputation on a daily air pollution data set with relatively small gaps. Our main contributions are: (1) an air pollution data set composed by several air pollution concentrations including criteria gases and thirteen meteorological covariates; and (2) a custom algorithm for data imputation of daily ozone concentrations based on a trend surface and a Gaussian Process. Data Visualization techniques were extensively used along this work, as they are useful tools for understanding the multi-dimensionality of point-referenced sensor data.

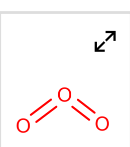
SciVal Topic Prominence

Topic: Kriging | Gaussian process | Spatial covariance

Prominence percentile: 92.413

Chemistry database information

Substances



Author keywords

Air pollution | Data imputation | Gaussian process | Sensor data

Indexed keywords

Metrics

PlumX Metrics

Usage, Captures, Mentions, Social Media and Citations beyond Scopus.

Related documents

Spatio-Temporal Poverty Analysis with INLA in Hierarchical Bayes Ecological Regression

Rachmawati, R.N. , Pusponegoro, N.H.
(2018) Procedia Computer Science

Hierarchical spatially varying coefficient and temporal dynamic process models using spTDyn

Bakar, K.S. , Kokic, P. , Jin, H.
(2016) Journal of Statistical Computation and Simulation

The application of a hierarchical bayesian spatiotemporal model for forecasting the saa trapped particle flux distribution

Suparta, W. , Gusrizal
(2014) Journal of Earth System Science

View all related documents based on references

Find more related documents in Scopus based on:

Authors Keywords

Engineering controlled terms: Climate change | Costs | Data visualization | Decision making | Gaussian distribution | Gaussian noise (electronic) | Ozone

Engineering uncontrolled terms: Data imputation | Environmental problems | Gaussian Processes | High resolution data | High resolution observations | Pollution concentration | Sensor data | Visualization technique

Engineering main heading: Air pollution

Funding details

Funding sponsor	Funding number	Acronym
U.S. Environmental Protection Agency See opportunities↗	EPA,EPA,EPA,EPA	
U.S. Environmental Protection Agency See opportunities↗	EPA,EPA,EPA,EPA	

Funding text #1

The Air Data project from the U.S. Environmental Protection Agency (EPA) collects air quality and weather measurements from more than 4000 outdoor monitors across the United States, Puerto Rico, and the Virgin Islands [22]. The data provided cover a time period from 1980 to 2017 in hourly, daily and annual aggregation. For this study, we downloaded the pre-generated daily files for the year 20162. The available files are grouped in two categories: (1) Criteria Gases such as Ozone, Sulfur dioxide (SO2), Carbon monoxide (CO), and Nitrogen dioxide (NO2); and (2) Particulates, such as PM2.5 (particles with a diameter of 2.5 μm or less), PM10 (particles with a diameter of 10 μm or less) [22].

Funding text #2

Motivated by the general challenge of studying spatio-temporal modeling of air pollution concentrations measured from cheap sensor networks, we decided to assemble a real-life data set made up of ozone concentrations taken from the Air Data project from the U.S. Environmental Protection Agency (EPA) plus weather covariates taken from the NCEP/NCAR Reanalysis Project (NNRP). The biggest problem with this data set, was the missing values in the target variable (ozone). This problem was addressed by means of a tailored algorithm inspired by the hierarchical models. Throughout this study we extensively applied several Exploratory Data Analysis techniques and multivariate analysis to understand the data and the relationships between the features.

ISSN: 21945357
ISBN: 978-303002827-5
Source Type: Book series
Original language: English

DOI: 10.1007/978-3-030-02828-2_27
Document Type: Conference Paper
Volume Editors: Botto-Tobar M.,Barba-Maggi L.,Villacres-Cevallos P.,Uvidia-Fassler M.I.,Gonzalez-Huerta J.,S. Gomez O.
Publisher: Springer Verlag

References (26)

View in search results format >

All | Export | Print | E-mail | Save to PDF | Create bibliography

1 Allenby, G.M., Rossi, P.E., McCulloch, R.E.
(2005) *Hierarchical Bayes Models: A Practitioners Guide*. Cited 20 times.

2 Bakar, K.S., Sahu, S.K.
spTimer: Spatio-temporal bayesian modeling using R (Open Access)

(2015) *Journal of Statistical Software*, 63 (15), pp. 1-32. Cited 39 times.
http://www.jstatsoft.org/v63/i15/paper
doi: 10.18637/jss.v063.i15

View at Publisher

3 Burke, J.A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.B.
(2006) *Participatory Sensing*. Cited 596 times.
Center for Embedded Network Sensing

4 Cameletti, M., Lindgren, F., Simpson, D., Rue, H.
Spatio-temporal modeling of particulate matter concentration through the SPDE approach

(2013) *AStA Advances in Statistical Analysis*, 97 (2), pp. 109-131. Cited 124 times.
doi: 10.1007/s10182-012-0196-3

View at Publisher

5 Campozano, L., Sánchez, E., Avilés, A., Samaniego, E.
Evaluation of infilling methods for time series of daily precipitation and temperature: The case of the ecuadorian andes
(2014) *Maskana*, 5 (1), pp. 99-115. Cited 22 times.

6 Cressie, N., Wikle, C.K.
(2015) *Statistics for Spatio-Temporal Data*. Cited 944 times.
Wiley, New York

7 Finley, A.O., Banerjee, S., Gelfand, A.E.
(2013) *Spbayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models*. Cited 7 times.

8 Gelfand, A.E.
Hierarchical modeling for spatial data problems

(2012) *Spatial Statistics*, 1, pp. 30-39. Cited 25 times.
doi: 10.1016/j.spasta.2012.02.005

View at Publisher

9 Gräler, B., Pebesma, E., Heuvelink, G.
Spatio-temporal interpolation using gstat

(2016) *R Journal*, 8 (1), pp. 204-218. Cited 90 times.
https://journal.r-project.org/archive/2016-1/na-pebesma-heuvelink.pdf

View at Publisher

10 Hasenfratz, D., Saukh, O., Sturzenegger, S., Thiele, L.
Participatory air pollution monitoring using smartphones
(2012) *Mob Sens*, 1, pp. 1-5. Cited 114 times.

11 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., (...), Joseph, D.
The NCEP/NCAR 40-year reanalysis project

(1996) *Bulletin of the American Meteorological Society*, 77 (3), pp. 437-471. Cited 19828 times.
http://ams.allenpress.com
doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

View at Publisher

12 Mukhopadhyay, S., Sahu, S.K.
A Bayesian spatiotemporal model to estimate long-term exposure to outdoor air pollution at coarser administrative geographies in England and Wales (Open Access)

(2018) *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 181 (2), pp. 465-486. Cited 5 times.
http://www.ingenta.com/journals/browse/bpl/rssa
doi: 10.1111/rssa.12299

View at Publisher

13 Pirani, M., Gulliver, J., Fuller, G.W., Blangiardo, M.
Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas (Open Access)

(2014) *Journal of Exposure Science and Environmental Epidemiology*, 24 (3), pp. 319-327. Cited 13 times.
doi: 10.1038/jes.2013.85

View at Publisher

14 (2013) *R: A Language and Environment for Statistical Computing*. Cited 62690 times.
https://www.r-project.org/

15 (2012) *Matrix Factorization as Data Imputation | S3l*
http://s3l.stanford.edu/blog/?p=66

16 Kumar, S.S., Bakar, K.S.
Hierarchical Bayesian auto-regressive models for large space time data with applications to ozone concentration modelling

(2012) *Applied Stochastic Models in Business and Industry*, 28 (5), pp. 418-419. Cited 8 times.
doi: 10.1002/asmb.1952

View at Publisher

17 Sahu, S.K., Gelfand, A.E., Holland, D.M.
High-resolution space-time ozone modeling for assessing trends

(2007) *Journal of the American Statistical Association*, 102 (480), pp. 1221-1234. Cited 70 times.
doi: 10.1198/016214507000000031

View at Publisher

18 Samworth, R.J.
Optimal weighted nearest neighbour classifiers (Open Access)

(2012) *Annals of Statistics*, 40 (5), pp. 2733-2763. Cited 72 times.
http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?handle=euclid.aos/1359987536&view=body&content-type=pdfview_1
doi: 10.1214/12-AOS1049

View at Publisher

19 Seo, J., Youn, D., Kim, J.Y., Lee, H.
Extensive spatiotemporal analyses of surface ozone and related meteorological variables in South Korea for the period 1999-2010 (Open Access)

(2014) *Atmospheric Chemistry and Physics*, 14 (12), pp. 6395-6415. Cited 18 times.
http://www.atmos-chem-phys.net/volumes_and_issues.html
doi: 10.5194/acp-14-6395-2014

View at Publisher

20 Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S.W., Shelow, D., (...), Preuss, P.W.
The changing paradigm of air pollution monitoring

(2013) *Environmental Science and Technology*, 47 (20), pp. 11369-11377. Cited 233 times.
doi: 10.1021/es4022602

View at Publisher

21 Stocker, M., Baranizadeh, E., Portin, H., Komppula, M., Rönkkö, M., Hamed, A., Virtanen, A., (...), Kolehmainen, M.
Representing situational knowledge acquired from sensor data for atmospheric phenomena

(2014) *Environmental Modelling and Software*, 58, pp. 27-47. Cited 16 times.
www.elsevier.com/inca/publications/store/4/2/2/9/2/1
doi: 10.1016/j.envsoft.2014.04.006

View at Publisher

22 Us, E.P.A.
(2016) *Air Data Basic Information | Air Data: Air Quality Data Collected at Outdoor Monitors across the US | US EPA*
https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information

23 Wen, H., Xiao, Z., Markham, A., Trigoni, N.
Accuracy Estimation for Sensor Systems

(2015) *IEEE Transactions on Mobile Computing*, 14 (7), art. no. 6898842, pp. 1330-1343. Cited 10 times.
http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?puNumber=7755
doi: 10.1109/TMC.2014.2352262

View at Publisher

24 (2016) *WHO Global Urban Ambient Air Pollution Database*. Cited 119 times.
http://www.who.int/phe/healthtopics/outdoorair/databases/cities/en/

25 Yanosky, J.D., Paciorek, C.J., Laden, F., Hart, J.E., Puett, R.C., Liao, D., Suh, H.H.
Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors (Open Access)

(2014) *Environmental Health: A Global Access Science Source*, 13 (1), art. no. 63. Cited 65 times.
http://www.ehjournal.net/home/
doi: 10.1186/1476-069X-13-63

View at Publisher

26 Zakaria, N.A., Noor, N.M.
Imputation methods for filling missing data in urban air pollution data for Malaysia

(2018) *Urbanism. Architecture. Constructions*, 9 (2), pp. 159-166.
http://uac.incd.ro/Art/v9n2a04.pdf

Gualán, R.; Department of Computer Science, University of Cuenca, Cuenca, Ecuador;
email:ronald.gualan@ucuenca.edu.ec