

# README for “Data driven approaches for passenger train delay estimation”.

Ren Wang and Daniel B. Work

May 4, 2015

## Abstract

This document describes the implementation of regression models introduced in the article “Data driven approaches for passenger train delay estimation” by Wang and Work, submitted to the IEEE ITSC 2015. The source code is hosted at [https://github.com/Lab-Work/TrainDelayEstimation\\_IIEEEITSC](https://github.com/Lab-Work/TrainDelayEstimation_IIEEEITSC).

## 1 License

This software is licensed under the *University of Illinois/NCSA Open Source License*:

Copyright (c) 2015 The Board of Trustees of the University of Illinois. All rights reserved.

Developed by: Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign  
[https://github.com/Lab-Work/TrainDelayEstimation\\_IIEEEITSC](https://github.com/Lab-Work/TrainDelayEstimation_IIEEEITSC)

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal with the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions: Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimers. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimers in the documentation and/or other materials provided with the distribution. Neither the names of the Department of Civil and Environmental Engineering, the University of Illinois at Urbana-Champaign, nor the names of its contributors may be used to endorse or promote products derived from this Software without specific prior written permission.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE CONTRIBUTORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS WITH THE SOFTWARE.

## 2 Running the code

Go to fold PythonCode, the provided .py-files can be used to reproduce the results presented in the publication.

1. Download the 2011-2013 Amtrak data from the link provided in reference [2] in the article.
2. Unzip the folders and copy them into the fold: amtrakData
3. run unzip.py to unzip all the data from 2011 to 2013
4. run delay\_extraction.py to compute the delay for all trains
5. run generate\_training\_data.py to generate training data
6. run main\_hist\_online\_68 to generate all figures shown in the paper, the results can be found under the folder named results.
7. run main\_online\_interacting.py to see the results for the online regression interacting model (3) and its comparison with online regression model (7) presented in the paper.