Researcher Dictionary
for
Time-Series Data

Version 1.0

September 17, 2014

REVISION HISTORY

Version	Date	Comments
1.0	09.17.2014	First version of document

INTRODUCTION

The following data dictionary describes the time-series variables available in the naturalistic driving data for use by the research community. In addition to this Introduction, this data dictionary includes six parts:

Revision History –This data dictionary should be considered a working document that will evolve over time. The revision history shown on the previous page provides a table which describes updates to the document.

Related Reading – A list of related subject areas and specific documents of value to users of the data set described in this data dictionary.

Data Description and Format – This section describes what data is available and how the data are stored.

List of Dictionary Fields – A description of the components or fields described in the dictionary for each variable entry.

Conversions, Coordinate System and Formulas – A catalog of unit conversions, sign conventions and formulas which may be of value to researchers working with these data.

List of Variables – A list of the entries (variables) in the dictionary which can be used as a table of contents to locate specific variables in the document.

Data Dictionary Entries – The dictionary entries themselves, one for each variable included in the data set.

RELATED READING

Individuals working with these data are encouraged to become familiar with them, the method in which they were collected, and literature in the area of secondary data analyses. The following references are provided as starting points to assist the researcher in his or her efforts.

Drowsy Driver Warning System Field Operational Test Overview

A Drowsy Driver Warning System (DDWS) detects physiological and/or performance indications of driver drowsiness and provides feedback to drivers regarding their state. The primary function of a DDWS is to provide information that will alert drivers to their drowsy state and motivate them to seek rest or take other corrective steps to increase alertness. The system tested in this study was the Driver Fatigue Monitor (DFM) developed by Attention Technologies, Inc., which estimates PERCLOS (percent eye closure). The primary goal of this field operational test (FOT) was to determine the safety benefits and operational capabilities, limitations, and characteristics of the DFM. The FOT was conducted in a naturalistic driving environment and data were collected from actual truck drivers driving commercial trucks. During the course of the study, 46 trucks were instrumented with a Data Acquisition System (DAS). Over 100 data variables such as the PERCLOS output from the DFM and driving performance data (e.g., lane position, speed, and longitudinal acceleration) were collected. Other collected measures included video, actigraphy, and questionnaires. The FOT had 103 drivers participate. Drivers were randomly assigned to either control (24 drivers) or experimental groups (79 drivers). The data collected include the following: approximately 46,000 driving-data hours; 397 load history files from 103 drivers; approximately 195,000 hours of activity/sleep data; questionnaires from all drivers; fleet management surveys from each company; and focus group results collected from 14 drivers during two post-study focus group sessions.

Methods

DDWS FOT

The methods used for collecting the DDWS FOT data are described in:

Hanowski, R.J., Blanco, M., Nakata, A., Hickman, J.S., Schaudt, W.A., Fumero, M.C., Olson, R.L., Jermeland, J., Greening, M., Holbrook, G.T., Knipling, R.R., & Madison, P. (September, 2008). *The drowsy driver warning system field operational test, data collection methods final report*. Report No. DOT HS 810 035. National Highway Traffic Safety Administration, USDOT, Washington, DC.

Knowledge Discovery in Data (KDD)

KDD is the general term to describe the larger process that includes what many people think of as data mining. The following references provide a useful introduction to the KDD process including topics such as data preparation, data transformations, evaluation of data mining approaches, etc.

Larose, Daniel T. (2005). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons, Inc. Hoboken, NJ.

Maimon, O., Rokach, Lior. Eds. (2005). Data mining and knowledge discovery handbook. Springer Science+Business Media, Inc. New York, NY.

Secondary Data Analysis

Use of data collected by other organizations is becoming increasingly common in this digital age. In some fields, such as the social sciences or business, the use of previously collected data is more common than, for example, in psychology or product development. The primary benefit of this approach is cost savings. There are also risks that can threaten the validity of analyses conducted in this manner. The following references include discussion and recommendations for secondary analysts.

Akerstrom, M., Jacobsson, K., Wasterfors, D. (2004). "Reanalysis of previously collected material" in Clive Seale, Giampietro Gobo, Jaber Gubrium, and David Silverman (eds), *Qualitative Research Practice*, Thousand Oaks, CA. Sage Publications Ltd.

Corti, L. Thompson, P. (2004). "Secondary analysis of archived data", in Clive Seale, Giampietro Gobo, Jaber Gubrium, and David Silverman (eds), *Qualitative Research Practice*, Thousand Oaks, CA. Sage Publications Ltd.

Dale, A. Arber, S., and Procter, M. (1988). *Doing Secondary Analysis*, Unwin Hyman Ltd., London.

Hyman, H. (1972). *Secondary Analysis of Sample Surveys*, Wesleyan University Press, Middletown, Connecticut.

Kiecolt, K. and Nathan, L. (1985). *Secondary Analysis of Survey Data – Sage University Paper Series on Quantitative Applications in the Social Sciences*, 53. Sage Publications, Beverly Hills, CA.

DATA DESCRITPION AND FORMAT

These data describe crashes or near-crashes from the DDWS FOT. The data include 16 crashes and 135 near-crashes. Each file contains the time series data spanning approximately 29s before the beginning of the event and 1s after event onset. The data in the files is stored as comma delimited text, with each column representing a variable, and each row representing a time sample.

LIST OF DICTIONARY FIELDS

For each of the variables, the dictionary provides the following seven fields:

1. # – A number used for referencing the rows in the dictionary.
2. Variable – A brief name for the variable

3. Column – The column in which the data for this variable are found in the text file
4. Units – The units in which the data is stored
5. Approximate Data Rate – The approximate rate at which the sensor recorded data
6. Sign Convention and Coding – the sign convention used for reporting the variable, or coding of variable
7. Notes – Notes helpful for using the variable appropriately

CONVERSIONS, COORDINATE SYSTEM AND FORMULAS

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

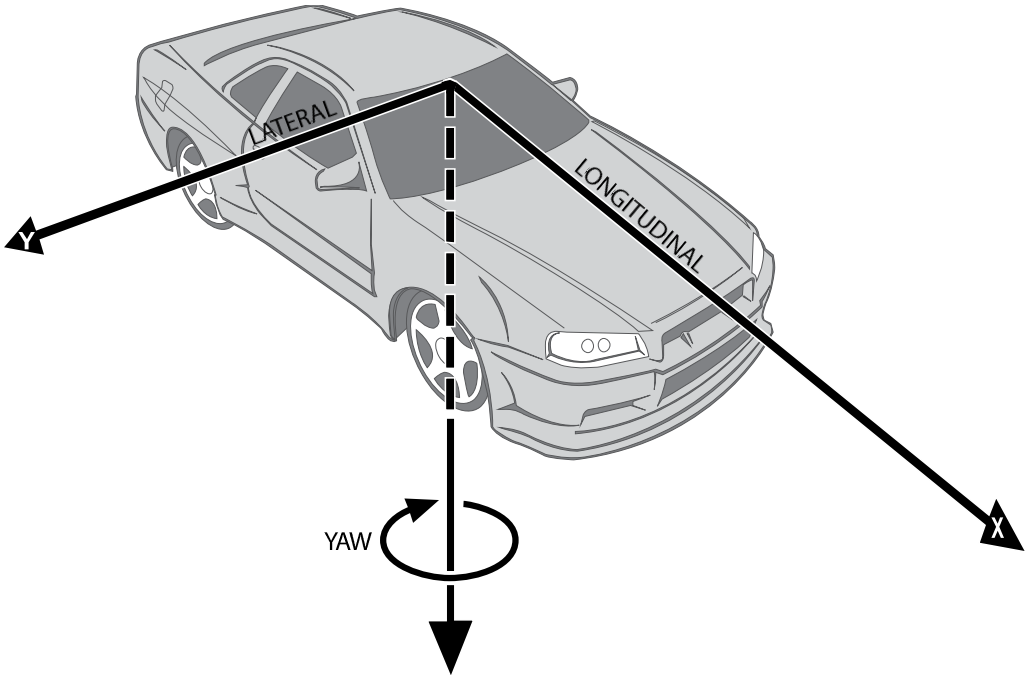

Figure 1. Coordinate System.

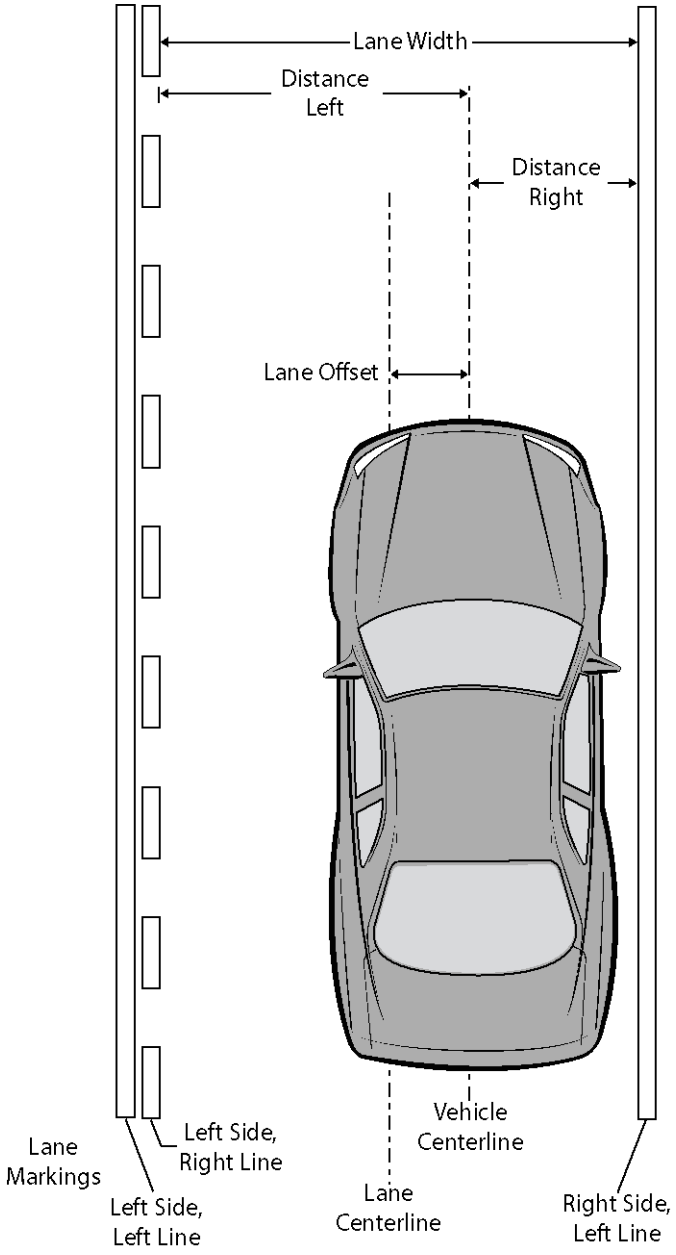

Figure 2. Machine-vision-based Lane Tracker Variable Names and Dimensions

LIST OF VARIABLES

The following variables are included in the text files.

Var	Name
1	event web id
2	Sync
3	Time
4	Gas pedal position
5	Speed, vehicle network
6	Yaw rate
7	Lateral acceleration
8	Longitudinal acceleration
9	Lane Markings, Continuity, Left Side Left Line
10	Lane Markings, Continuity, Left Side, Right Line
11	Lane Markings, Continuity, Right Side Left Line
12	Lane Markings, Continuity, Right Side, Right Line
13	Lane Markings, distance left
14	Lane Markings, distance right
15	Lane Markings, type left
16	Lane Markings, type right
17	Lane markings, probability left
18	Lane markings, probability right
19	Radar, forward, ID
20	Radar, forward, range
21	Radar, forward, range rate
22	Radar, forward azimuth
23	Brake on off
24	Turn signal - right
25	Turn signal - left

Data Dictionary Entries

#	Variable	Column	Units	Approx. Data Rate	Sign Convention and Coding	Notes
1	Trip Identifier	1	-	-		
2	Sync	2	-	10 Hz		Increasing integer for each row of data within a file
3	Time	3	ms	10 Hz		
4	Gas pedal position	4	-	3-10 Hz	Increasing value indicates increasing deflection.	
5	Speed, vehicle network	5	km/h	3-10 Hz	Forward and reverse motion is positive	
6	Yaw rate	6	deg/s	10 Hz	Positive if vehicle turns to right.	
7	Lateral acceleration	7	g	10 Hz	Positive indicates lateral acceleration as generated by the vehicle turning to right.	
8	Longitudinal acceleration	8	g	10 Hz	Positive indicates longitudinal acceleration as generated by the vehicle accelerating from a stop.	
9	Lane Markings, Continuity, Left Side, Left Line	9	-	10 Hz	0=solid, 1=dash, 2=unsure	
10	Lane Markings, Continuity, Left Side, Right Line	10	-	10 Hz	0=solid, 1=dash, 2=unsure	
11	Lane Markings, Continuity, Right Side, Left Line	11	-	10 Hz	0=solid, 1=dash, 2=unsure	
12	Lane Markings, Continuity, Right Side, Right Line	12	-	10 Hz	0=solid, 1=dash, 2=unsure	

#	Variable	Column	Units	Approx. Data Rate	Sign Convention and Coding	Notes
13	Lane Markings, distance left	13	in	10 Hz	Negative normal condition. Movement left in lane increases value toward zero. Positive during camera centerline crossing left marker.	
14	Lane Markings, distance right	14	in	10 Hz	Positive normal condition. Movement right in lane reduces value toward zero. Negative when camera center line crossing right marker.	
15	Lane Markings, type left	15	-	10 Hz	0 = none, 1 = double line, 2 = single line, 3 = road gutter, 4 = road edge	
16	Lane Markings, type right	16	-	10 Hz	0 = none, 1 = double line, 2 = single line, 3 = road gutter, 4 = road edge	
17	Lane markings, probability left	17	-	10 Hz	Probability that machine vision lane marking system is providing correct data for the left side lane markings. Higher values indicate more probable.	
18	Lane markings, probability right	18	-	10 Hz	Probability that machine vision lane marking system is providing correct data for the right side lane markings. Higher values indicate more probable.	
19	Radar, forward, ID	19-25	-	10 Hz	Cycles 1 thru 255 as new targets are identified.	Target ID provided for seven potential physical targets. Over time, same target may appear in any or none of seven columns.

#	Variable	Column	Units	Approx. Data Rate	Sign Convention and Coding	Notes
20	Radar, forward, range	26-32	ft	10 Hz		Range to seven potential targets. Use Radar, forward ID columns to identify which column to query for a given target's range.
21	Radar, forward, range rate	33-39	ft/s	10 Hz	Positive values indicate distance to target increasing	Range rate to seven potential targets. Use Radar, forward ID columns to identify which column to query for a given target's range rate.
22	Radar, forward azimuth	40-46	rads	10 Hz	Positive value to right (passenger side) of forward facing radar center line.	Azimuth to seven potential targets. Use Radar, forward ID columns to identify which column to query for a given target's azimuth.
23	Brake on off	47	-	3-10 Hz	0 = off, 1 = on	
24	Turn signal right	48	-	3-10 Hz	0 = off, 1 = on	
25	Turn signal left	49		3-10 Hz	0 = off, 1 = on	