Researcher Dictionary for Time-Series Data

Version 1.2

September 7, 2010

Revision History

Version	**Date**	**Comments**
1.0	11/3/2008	First version of document
1.1	3/4/2009	Adjusted column values
1.2	9/7/2010	Added explanation for composite speed

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.

Description of the Data 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.

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.

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.

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.

100-Car Study Overview

The 100-Car Naturalistic Driving Study was an instrumented vehicle study conducted in the Northern Virginia / Washington, D.C. area over a two year period. The primary purpose of the study was to collect large scale naturalistic driving data. To this end the instrumentation was designed to be unobtrusive, study participants were given no special instructions, and experimenters were not present. Approximately 100 vehicles were instrumented with a suite of sensors including forward and rearward radar, lateral and longitudinal accelerometers, gyro, GPS, access to the vehicle CAN, and five channels of compressed digital video. Collection rates for the various sensors ranged from 1Hz to 10Hz. This collection effort resulted in approximately 2,000,000 vehicles miles and 43,000 hours of driving data.

Methods

<u>100-Car Methods</u>

The methods used for collecting the 100-car data are described in:

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R. R.(2006) The 100-Car Naturalistic Driving Study, Phase II - Results of the 100-Car Field Experiment DOT HS 810 593.

<u>Knowledge Discovery in Data (KDD)</u>

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.

<u>Secondary Data Analysis</u>

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.

Description of the Data and Format

These data describe crashes or near-crashes from the 100-Car Naturalistic Driving Study (Dingus et al., 2006). The data include 68 crashes and 760 near-crashes. Each file contains the time series data spanning 30s before an event and 10s after an event.   The data in the files is stored as comma delimited text, with each column representing a variable, and each row representing a time sample.

Event Descriptions

Summary narratives that provide situational context for each crash and near crash event are provided in a pdf document, entitled 100CarEventNarrative.pdf, located with the reduced event dataset. Detailed information about event parameters, environmental conditions, and driver states during these events can be found in the reduced video event dataset. A separate researcher dictionary is provided for the reduced video event dataset.

Sensor Status

For each crash and near crash a series of diagnostic evaluations on the time series data were utilized to determine sensor operation status.   This sensor operational status information is provided in an associated pdf document entitled EventSensorStatus.pdf.

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)

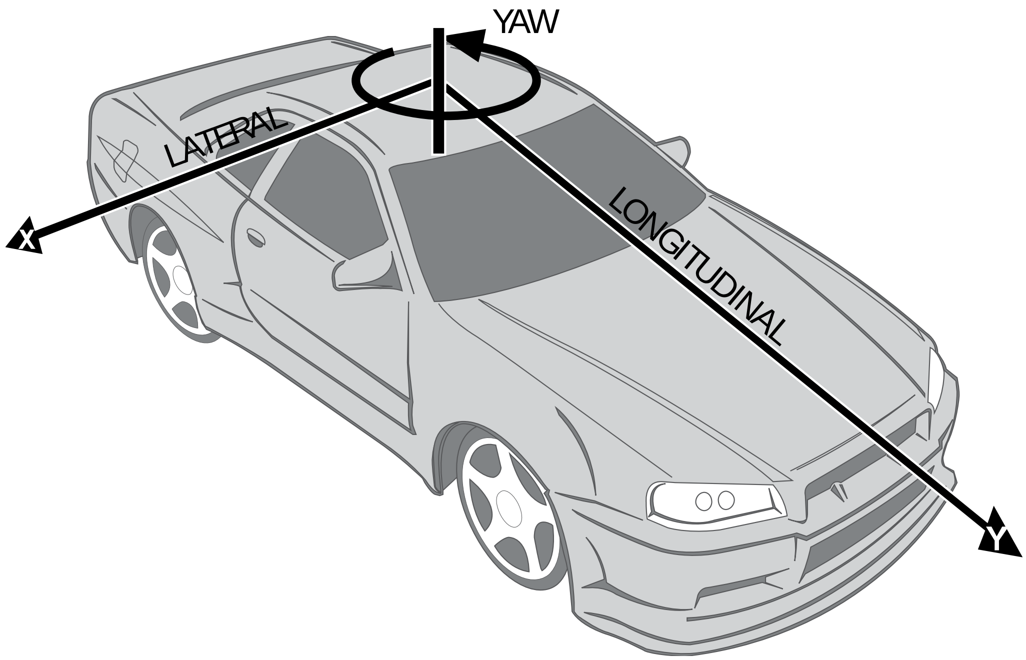

Figure 1. Coordinate System (Note not to SAE standard).

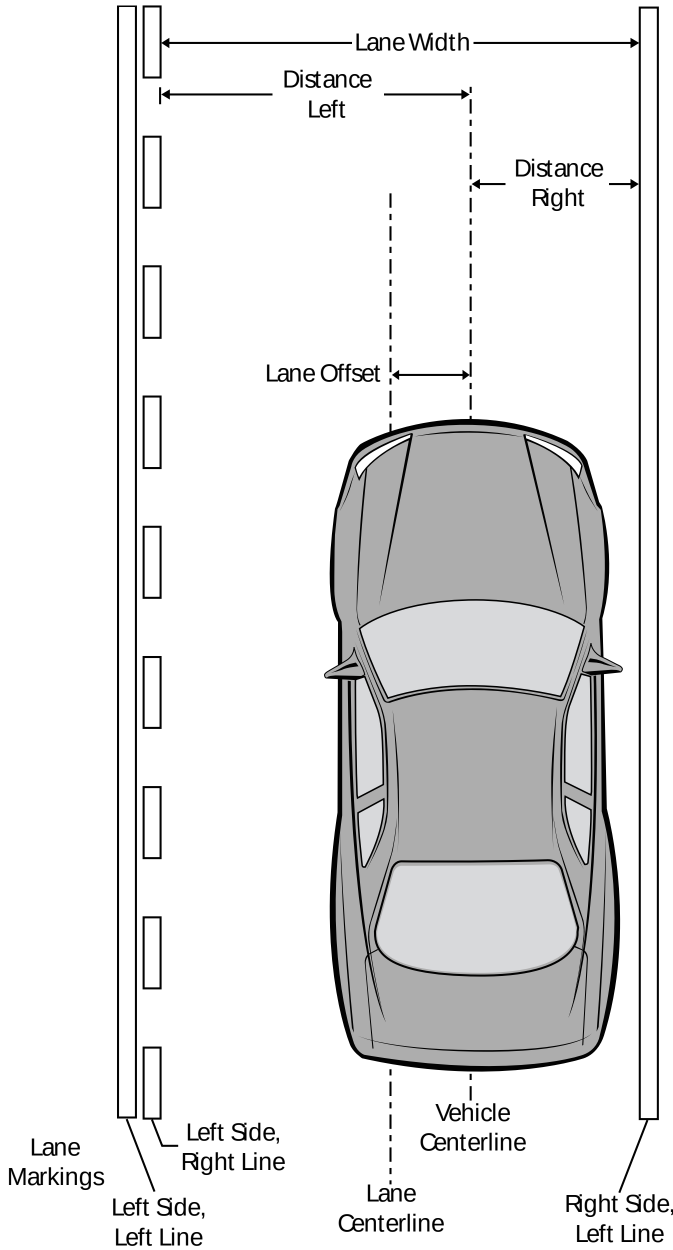

Figure 2. Machine-vision-based Lane Tracker Variable Names and Dimensions

List of Variables

The following variables are included in the text files.

Var	Name
1	Trip Identifier
2	Sync
3	Time
4	Gas pedal position
5	Speed, vehicle composite
6	Speed, GPS horizontal
7	Yaw rate
8	Heading, GPS
9	Lateral acceleration
10	Longitudinal acceleration
11	Lane Markings, Continuity, Left Side Left Line
12	Lane Markings, Continuity, Left Side Right Line
13	Lane Markings, Continuity, Right Side, Left Line
14	Lane Markings, Continuity, Right Side, Right Line
15	Lane Markings, distance left
16	Lane Markings, distance right
17	Lane Markings, type left
18	Lane Markings, type right
19	Lane markings, probability left
20	Lane markings, probability right
21	Radar, forward, ID
22	Radar, rearward, ID
23	Radar, forward, range
24	Radar, rearward, range
25	Radar, forward, range rate
26	Radar, rearward, range rate
27	Radar, forward azimuth
28	Radar, rearward azimuth
29	Light intensity
30	Brake on off
31	Turn signal state

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	s	10 Hz		
4	Gas pedal position	4	-	3-10 Hz	Increasing value indicates increasing deflection. Only appropriate as delivered for detecting trend of input within one file. Scale not equal across vehicles and/or trips. Sensor not necessarily at zero when no input is present.	
5	Speed, vehicle composite	5	mph	3-10 Hz	Forward and reverse motion is positive	-1 is used to denote instances in which composite speed cannot be determined.
6	Speed, GPS horizontal	6	mph	1 Hz	Forward and reverse motion is positive	
7	Yaw rate	7	deg/s	10 Hz	Positive in vehicle turns to left.	
8	Heading, GPS	8	deg	1 Hz	0-359, 0=North, 90=East, 180=South, 270=West	
9	Lateral acceleration	9	g	10 Hz	Positive indicates lateral acceleration as generated by the vehicle turning to left.	X is lateral in this data set (non standard)
10	Longitudinal acceleration	10	g	10 Hz	Positive indicates longitudinal acceleration as generated by the vehicle accelerating from a stop.	Y is longitudinal in this data set (non standard)

#	Variable	Column	Units	Approx. Data Rate	Sign Convention and Coding	Notes
11	Lane Markings, Continuity, Left Side Left Line	11	-	10 Hz	0=solid, 1=dash, 2=unsure	
12	Lane Markings, Continuity, Left Side Right Line	12	-	10 Hz	0=solid, 1=dash, 2=unsure	
13	Lane Markings, Continuity, Right Side, Left Line	13	-	10 Hz	0=solid, 1=dash, 2=unsure	
14	Lane Markings, Continuity, Right Side, Right Line	14	-	10 Hz	0=solid, 1=dash, 2=unsure	
15	Lane Markings, distance left	15	in	10 Hz	Negative normal condition. Movement left in lane increases value toward zero. Positive during camera centerline crossing left marker.	
16	Lane Markings, distance right	16	in	10 Hz	Positive normal condition. Movement right in lane reduces value toward zero. Negative when camera center line crossing right marker.	
17	Lane Markings, type left	17	-	10 Hz	0 = none, 1 = double line, 2 = single line, 3 = road gutter, 4 = road edge, 5 = reflector	
18	Lane Markings, type left	18	-	10 Hz	0 = none, 1 = double line, 2 = single line, 3 = road gutter, 4 = road edge, 5 = reflector	
19	Lane markings, probability left	19	-	10 Hz	0 = none, 1 = double line, 2 = single line, 3 = road gutter, 4 = road edge, 5 = reflector	
20	Lane markings, probability right	20	-	10 Hz	0 = none, 1 = double line, 2 = single line, 3 = road gutter, 4 = road edge, 5 = reflector	

#	Variable	Column	Units	Approx. Data Rate	Sign Convention and Coding	Notes
21	Radar, forward, ID	21-27	-	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.
22	Radar, rearward, ID	28-34	-	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.
23	Radar, forward, range	35-41	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.
24	Radar, rearward, range	42-48	ft	10 Hz		Range to seven potential targets. Use Radar, rearward ID columns to identify which column to query for a given target's range.
25	Radar, forward, range rate	49-55	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.
26	Radar, rearward, range rate	56-62	ft/s	10 Hz	Positive values indicate distance to target increasing	Range to seven potential targets. Use Radar, rearward ID columns to identify which column to query for a given target's range rate.
27	Radar, forward azimuth	63-69	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.

#	Variable	Column	Units	Approx. Data Rate	Sign Convention and Coding	Notes
28	Radar, rearward azimuth	70-76	rads	10 Hz	Positive value to right (driver side) of rearward facing radar center line.	Azimuth to seven potential targets. Use Radar, rearward ID columns to identify which column to query for a given target's azimuth.
29	Light intensity	77	-	10 Hz	Increasing value indicates increasing light intensity from either natural or manufactured sources (e.g. headlamps, overhead lighting). Only appropriate as delivered for detecting trend in light intensity within one file at night. Scale not equal across vehicles and/or trips.	
30	Brake on off	78	-	3-10 Hz	0 = off, 1 = on	
31	Turn signal state	79	-	3-10 Hz	0 = off, 1 = left, 2 = right, 3 = both	