The Vehicle Data Translator V3.0 System Description

www.its.dot.gov/index.htm
Final Report — May 30, 2011
FHWA-JPO-11-127

image courtesy iStockPhoto (http://www.istockphoto.com/stock-photo-11156851-rush-hour-traffic-in-the-rain.php)

U.S. Department of Transportation
Federal Highway Administration
Research and Innovative Technology Administration

Produced by DTFH61-08-D-00012
ITS Joint Program Office
Research and Innovative Technology Administration
U.S. Department of Transportation

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

Technical Report Documentation Page

1. Report No. **FHWA-JPO-11-127**	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle The Vehicle Data Translator V3.0 System Description		**5. Report Date** May 30, 2011 **6. Performing Organization Code**
7. Author(s) Sheldon Drobot, Michael Chapman, Brice Lambi, Gerry Wiener, and Amanda Anderson		**8. Performing Organization Report No.**
9. Performing Organization Name And Address University Corporation for Atmospheric Research 3090 Center Green Dr. Boulder, CO 80301		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** DTFH61-08-D-00012
12. Sponsoring Agency Name and Address United States Department of Transportation Research and Innovative Technology Administration Federal Highway Administration, Office of Operations 1200 New Jersey Ave., SE Washington, DC 20590		**13. Type of Report and Period Covered** **14. Sponsoring Agency Code** FHWA, HOTO
15. Supplementary Notes Mr. Gabriel Guevara (COTM)		

16. Abstract

With funding and support from the USDOT RITA and direction from the FHWA Road Weather Management Program, NCAR is developing a Vehicle Data Translator (VDT) software system that incorporates vehicle-based measurements of the road and surrounding atmosphere with other weather data sources. This document describes version 3.0 of the VDT, including data ingest, data quality check, derivation of road and weather statistics, and the open-source philosophy.

17. Key Words Connected vehicles, road weather, probe data	**18. Distribution Statement** This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 45	**22. Price** **$0.00**

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Acknowledgements

This work is sponsored by the U.S. Department of Transportation Road Weather Management Program and the Intelligent Transportation Systems Joint Program Office under contract DTH61-08-D-00012.

Disclaimer Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation.

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

The United States Government does not endorse products or manufactures. Trade and manufactures' names appear in the report only because they are considered essential to the objective of the document.

The opinions, findings, and conclusions expressed in this document are those of the authors and not necessarily those of the U.S. Department of Transportation, Federal Highway Administration, or any other sponsor.

Table of Contents

List of Tables

List of Figures

Executive Summary

The U.S. Department of Transportation's (USDOT) Federal Highway Administration (FHWA) and Research and Innovative Technology Administration (RITA) are jointly working to promote safety, mobility, and the environment on the nation's surface transportation system through a new connected vehicle initiative. This initiative is a multimodal effort to enable wireless communications among vehicles, the infrastructure, and passengers' personal communication devices. It will enhance Americans' safety, mobility, and quality of life, while helping to reduce the environmental impact of surface transportation.

This report provides specific direction for continued work that the University Corporation for Atmospheric Research's (UCAR) National Center for Atmospheric Research (NCAR) started related to the Vehicle Data Translator (VDT) system that ingests and processes mobile data already resident on the vehicle along with ancillary weather data (e.g., radar).

This report begins by overviewing the three stages of the VDT. In Stage I, the VDT ingests mobile data and performs simple Quality-Control (QC) routines, such as ensuring that the mobile observations are located on the Earth and within an appropriate hour window (e.g., 0-24 hours). In Stage II, mobile observations are run through more rigorous QC and basic road segment data, such as the mean temperature on a road segment, are also generated. In the final stage, mobile and ancillary data (e.g., radar) are combined to create advanced road segment data, such as road slickness potential.

The report continues with brief comments on the types of data and data fornats that the VDT ingests, and then continues with a description of the ancillary data the VDT is designed to use. This latter list includes radar, gridded weather observations, satellite data, surface stations, National Weather Service products, and social media.

The report also outlines the VDT QC procedures, some of which closely resemble QC routines from the *Clarus* System, and others that are new to the mobile observation world. The former list includes the Anticipated Range, Time Step, Persistence, Spatial, and Climate Range Tests. The newer tests include the Data Filtering, Model Analysis, Neighboring Vehicle, and Combined Algorithm tests.

Following a brief description of the statistical routines used in the Basic Road Assessment output (Stage II), this report provides a detailed discussion on the method for derving the Advanced Road Segment Assessments (Stage III) for Precipitation, Pavement Condition, Visibility, and All-Hazards.

The report concludes by noting that the software suite emphasizes the use of open source software and avoids proprietary software packages unless such packages provide a significant advantage in initial development time or in establishing proof of concept.

Chapter 1 Introduction

The U.S. Department of Transportation's (USDOT) Federal Highway Administration (FHWA) and Research and Innovative Technology Administration (RITA) are jointly working to promote safety, mobility, and the environment on the nation's surface transportation system through a new connected vehicle initiative. This initiative is a multimodal effort to enable wireless communications among vehicles, the infrastructure, and passengers' personal communication devices. It will enhance Americans' safety, mobility, and quality of life, while helping to reduce the environmental impact of surface transportation.

This report provides specific direction for continued work that the University Corporation for Atmospheric Research's (UCAR) National Center for Atmospheric Research (NCAR). Namely, NCAR developed a Vehicle Data Translator (VDT) system that ingests and processes mobile data already resident on the vehicle along with ancillary weather data (e.g., radar). Continued work on the VDT revolves around several keys concepts:

- Developing the connected vehicles' role in the overall FHWA Road Weather Management Program (RWMP) goal of "Anytime, Anywhere Road Weather Information"
- Improving the characterization of current weather and road conditions, especially from a spatial perspective
- Exploiting any and all vehicle-based data, including:
    - Data from Original Equipment Manufacturer (OEM) sensors (e.g., air temperature, wiper status, braking status)
    - Data from after-market sensors (e.g., pavement temperature, plow status)
- Combining data from vehicles with data from fixed sources (e.g., *Clarus*)
- Outputting basic and inferred segment-based weather and road conditions to support weather-related applications

Earlier versions of the VDT were developed using 9 to 11 vehicles operating in the Detroit Development Test Environment (DTE) in the winter and spring of 2009 and 2010. Additional details on this work are presented in NCAR (2009a, 2010a). Using the vehicle data, as well as standard weather observations, e.g., radar, satellite, and model analyses, NCAR developed algorithms to derive road and atmospheric hazard condition assessments as part of the VDT 2.0.

The long-term FHWA goal for connected vehicles includes the collection of vehicular and environmental status data by millions of passenger and commercial vehicles. The data will be transmitted to other vehicles and to the infrastructure to be used for safety, mobility, and environmental applications. Environmental data will include weather and road condition data, such as ambient air temperature and atmospheric pressure, road surface temperature, and road friction coefficient. Some vehicle status data will also be related to weather and road conditions, such as windshield wiper status, antilock braking system (ABS) status, electronic stability, and traction control status. Ultimately, decision-makers will have the benefit of decision support tools that have access to data provided by millions of vehicles. However, the specific development of decision support tools is not a foundational platform of this project. Rather, here the FHWA RWMP desires to demonstrate how weather, road condition, and related vehicle data may be collected, transmitted, and processed. Using existing fleet infrastructures, data sets, sensors, and wireless communications technology provided by State Departments of Transportation (DOTs), this project will help to determine standards and procedures by prototyping the process of integrating weather, road condition, and vehicle status data messages.

This project builds on the capabilities of the VDT to ingest mobile weather, road condition, and vehicle status data, check the data quality, and aggregate the data for use in applications. The project also features the integration of mobile weather and road condition data into the FHWA's *Clarus* System. *Clarus*, operated by Mixon Hill Inc. under contract to the FHWA, currently organizes weather and road condition data collected by stationary sensors across the U.S. and parts of Canada, and makes the data available over the Internet with text and graphics-based retrieval systems.

The overall objectives of this project are to:

1. Determine requirements for collection and processing of weather, road condition, and vehicle status data from mobile sources.
2. Enhance the VDT tool and use it to perform quality checks and consolidate weather and road condition data.
3. Demonstrate the value of adding weather, road condition, and vehicle status data from mobile sources into management or decision support systems and other road weather-based applications.
4. Provide data input to the *Clarus* system, the Data Capture and Management program, and other programs.

Chapter 2 Task Objectives

The U.S. Department of Transportation's (USDOT) Federal Highway Administration (FHWA) and Research and Innovative Technology Administration (RITA) are jointly working to promote safety, mobility, and the environment on the nation's surface transportation system through a new connected vehicle initiative. This initiative is a multimodal effort to enable wireless communications among vehicles, the infrastructure, and passengers' personal communication devices. It will enhance Americans' safety, mobility, and quality of life, while helping to reduce the environmental impact of surface transportation.

This Task, entitled "Data Processing and Transmission of Processed Data", encompasses several goals aimed at developing the VDT 3.0, processing and analyzing collected data, and transmitting the data to the Clarus System. Specifically, this report outlines four topical areas, of which the last two form the bulk of this report:

- The VDT data ingest and processing routines, including the necessary metadata fields for the added observations
- The process to establish the data and the metadata fields for the collection of discrete (point-data) mobile data sources and for average (road segment) sources
- The VDT quality check algorithms, including the development of new confidence value flags
- The VDT algorithms for deriving new or enhanced weather, road condition, and environmental conditions from available data, for example by using windshield wiper status to infer precipitation or by using ABS and tire data to infer road friction. Upon FHWA concurrence, these new algorithms and products will be developed and tested.

Chapter 3 VDT 3.0 Data Processing and Comparison with VDT 2.0

This section outlines some basic terms, discusses the three main stages in the VDT 3.0, and provides some comments on VDT-related applications and data storage.

3.1. Basic Terminology and Input Data

The definition of several terms will aid the reader in understanding this document:

Input data are referred to as one of two things:

- Mobile data are all data originating from a vehicle, whether native to the Controller Access Network Bus (CANBus) or as an add-on sensor (e.g., pavement temperature sensor mounted to a vehicle).
- Ancillary data represent all other data, such as surface weather stations, model output, satellite data, and radar data.

Within the VDT, three data types are referenced:

- Point data are individual data points with a known time stamp and geographic location. An example would be a single air temperature reading from a moving vehicle.
- Gridded data represent a spatial area on the earth, such as a 5-km^2 box. Gridded data can consist of a single observation (e.g., radar pixel) or a combination of data points statistically compiled within a grid cell (e.g., average temperature in a grid cell from several surface stations).
- Segment data represent a given stretch of roadway with a single data value. This data may be an average of several point data values (e.g., average speed of five vehicles over a 1-mile segment), or the fusion of several point data sources (e.g., combining wiper, brake, and temperature data to determine road slickness potential).

The VDT also contains several data formats:

- Raw data are simply geolocated data.
- Quality-checked (QC) data are raw data after some form of QC has been performed.
- Basic road segment data are individual QC data compiled via some statistical procedure. Examples include the mean air temperature along a segment, or the number of wiper data points set to ‘off’. The basic road segment statistics do not combine different types of input data into a single product. For instance, a basic road segment output would never combine wipers and temperature data into a single road condition analysis.
- Advanced road segment data combine at least two types of QC’d data. An example is combining road temperature and brake information to generate a ‘road slickness’ assessment.

Finally, some terms are used in this document that may not be familiar to the reader, so Appendix A provides a List of Acronyms and Appendix B provides a Glossary.

3.2. VDT Overview

The organizational construct for how the data are processed in the VDT 3.0 is displayed in Figure 3-1.

3.2.1. Stage I – Mobile Data Collection, Parsing, and Sorting

The initial stage of the VDT ingests mobile data. If the data are already pre-processed in some way, such as by the *Clarus* System, then the VDT can simply read the metadata and data from the *Clarus* output. However, the VDT also has routines to directly ingest mobile data from the CANBus or a data collection and forwarding facility, parse them, and then sort them by time, road segments, and grid cells (the road segments and grid cells are user-defined via configuration files). This stage also reads any extra data sent from the vehicle that originates outside of the CANBus, such as readings from add-on sensors. These data are then passed through a *QC Module* that tags data that contain invalid geospatial or temporal information (e.g., latitude values greater than 90°N or time of day greater than 23:59:59). All data are passed through the *Output Data Handler*, which outputs the “parsed mobile data” for use in applications, and also for use in Stage II of the VDT. One such application has already been created, and it can display the data from this stage on a Google Map. More details on this application are listed in Section 3.3.

3.2.2. Stage II – Basic Road Segment Data

Stage II analyses provide the road segment data using QC’d mobile data. The *QC Module* examines individual mobile data (CANBus and some add-on sensors) and flags each data point for the relevant QC tests that are listed in Table 3-1 and outlined in Section 6. Ancillary data, such as *Clarus* surface station data and radar data, are also ingested by the *Ancillary Data Ingesters*, which perform the same functions as the Stage I *Mobile Data Ingestors module* (except in this case for ancillary data), including time stamping and geolocating. These ancillary data are then used in some of the QC processes, but they are not QC’ed themselves; however, the ancillary data used in the VDT 3.0 is all QC’d by other means before being incorporated into the VDT. All data are passed through to the *Statistics* component, where the mobile data that pass QC are used to compute road segment statistics. Examples of these data would include the mean air temperature over an individual road segment for a given time step, or the percentage of windshield wipers activated over an individual road segment for a given time step. All mobile data with QC flags and the statistical data for the “Basic road segment” data are output from Stage II.

3.2.3. Stage III – Advanced Road Segment Data

Stage III analyses provide additional value-adds for mobile data. In the *Inference Module*, fuzzy logic algorithms, decision trees, and other data mining procedures are used to produce the “Advanced road segment” data. Examples of these include combining mobile data with radar, satellite, and fixed surface station data to compute a derived ‘road precipitation’ product over an individual road segment for a given time step. These data are then run through a *QC Module* that assigns a confidence value to the “Advanced road segment” data assessments.

3.3. Apps and Other Data Environments

Data from the VDT will be used in applications and other data environments. Although not part of the VDT 3.0, NCAR has built one application, the Pathfinder VDT Display, which can show all data output from the VDT 3.0 (stages I, II, or III), as well as (a) National Weather Service (NWS) watches, warnings, and advisories; (b) NWS Storm Prediction Center (SPC) storm reports; (c) radar and satellite data; and (d) social media feeds, such as Twitter.

3.4. Data Storage

Data are not archived in the VDT 3.0, but data sent from the VDT to other data environments, such as *Clarus*, may be archived.

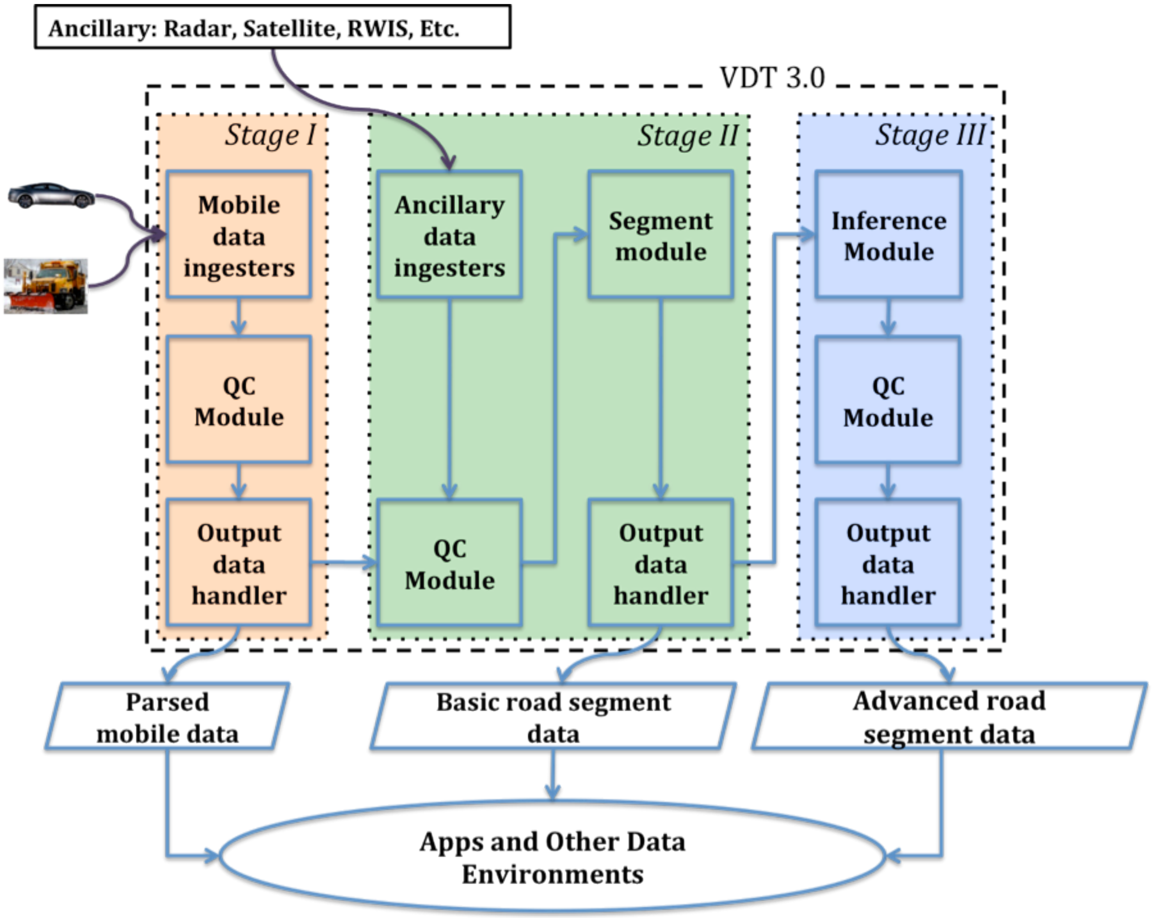


Figure 3-1: VDT 3.0 Data Processing. *Images courtesy iStockPhoto* (http://www.istockphoto.com/stock-photo-1285117-snow-plow.php and http://www.istockphoto.com/stock-photo-9272930-aluminium-supercar.php)

Table 3-1. Comparison of some data and process changes in VDT 3.0 versus VDT 2.0

Module / Process	VDT 2.0	VDT 3.0
Input Data		
Probe Data Elements	*External air temperature Atmospheric pressure Wiper status Headlight status Accelerometer Anti-lock braking status Traction control Stability control Rate of change of steering Vehicle velocity Brake status Brake boost Date Time Location Vehicle heading Rain (rain sensor)*	*External air temperature Atmospheric pressure Wiper status Headlight status Accelerometer Anti-lock braking status Traction control Stability control Rate of change of steering Vehicle velocity Brake status Brake boost Date Time Location Vehicle heading Rain (rain sensor) Relative humidity Pavement temperature RPM Engine torque Exhaust diagnostics*
Radar	*Composite reflectivity*	*Composite reflectivity Echo top Velocity Precipitation Vertically integrated liquid Dual-polarization data*
Gridded Weather Observations	*Rapid Update Cycle (RUC) Surface Assimilation Systems (RSAS)*	*RUC Real-Time Mesoscale Analysis (RTMA)*
Satellite	*Geostationary Operational Environmental Satellite (GOES): 2- and 4-km cloud mask (present / not present)*	*GOES: 2- and 4-km cloud mask (Naval Research Laboratory cloud classification algorithm)*
Surface stations	*Automated Surface Observing System (ASOS) / Automated Weather Observing System (AWOS)*	*ASOS/AWOS Clarus, MADIS*
Road Weather Forecasts	*Not present*	*NCAR Road Weather Forecast System (RWFS)*
NWS Products	*National Oceanic and Atmospheric Administration (NOAA) Storm Prediction*	*SPC Storm Reports NWS Watch / Warning / Advisory information*

	Center (SPC) Storm Reports *National Weather Service (NWS) Watch / Warning / Advisory information* *NWS Story of the Day*	*NWS Story of the Day*
Social Media	*Twitter (#wxreport)*	*Twitter (#wxreport, #STwx[1])*
Quality Check Routines		
	Clarus-based: *Anticipated Range Test* *Non-Clarus based:* *Data Filtering Test* *Neighboring Station Test* *Model Analysis Test* *Combined Algorithm Test* *Neighboring Vehicle Test*	*Clarus-based:* *Anticipated Range Test* *Persistence Test* *Step Test* *Spatial Tests (replaces Neighboring Station Test)* *Climate Range Test* *Non-Clarus based:* *Data Filtering Test* *Model Analysis Test (Enhanced)* *Neighboring Vehicle Test* *Combined Algorithm Test*
VDT Algorithms		
	Precipitation *Pavement Condition* *Visibility* *All-hazards*	*Precipitation (Enhanced)* *Pavement Condition (Enhanced)* *Visibility* *All-hazards (Enhanced)*

[1] ST stands for the 2-letter state modifier, such as #OKwx, #COwx, etc.

Chapter 4 Mobile Data Ingesters

In version 2.0 of the VDT, the mobile data ingesters were hard-coded to accept many different types of vehicle data files. The data would then be parsed and reformatted to either (1) a user-defined grid and time step; or (2) the default one-mile segment with a five-minute update step. For version 3.0, the parser will be modified to accept new vehicle data feeds and extra fields provided to the VDT from those new data feeds. The new vehicle data is expected from the Minnesota Department of Transportation (Mn/DOT) and the Nevada DOT (NDOT). If other data feeds become available [e.g., Strategic Highway Research Program 2 (SHRP2), Vehicle-based Information and Data Acquisition System (VIDAS), and/or Mobile Platform Environmental Data (MoPED)], the mobile data ingesters will be modified to accept those new formats accordingly. Additionally, in version 3.0, data can be parsed and reformatted to either (a) a user-defined grid and time step, (b) the default one-mile segment with a five-minute update step, or (c) a simple data point (i.e., not assigned to a grid or road segment).

Although the overall data requirements for mobile data are still in the development phase, some initial considerations on data requirements are presented in Appendix B. The VDT 3.0 will function best if the following data, at a minimum, are available:

Input data: External air temperature, wiper status, headlight status, ABS status, rate of change of steering wheel, vehicle velocity, date, time, location, vehicle heading, and pavement temperature. These must be collected at a minimum of once every five minutes, but optimally are collected every 20 seconds at present, with a future goal of once per second.

Ancillary data: composite reflectivity, satellite cloud mask, surface station data. These data require updating at a minimum of every five minutes.

The VDT requires data be in J2735, J1939, J1979, or Network Common Data Format (netCDF) formats.

Chapter 5 Ancillary Data Ingesters

Ancillary data play a key role in the VDT, serving both as a data field for QC of the mobile data and as inputs for the *Inference Module*. As stated previously, ancillary data are not QC'd. This section outlines the ancillary data used by VDT 3.0.

5.1. Radar

For the VDT 2.0, radar data are ingested from the National Oceanic and Atmospheric Administration's National Severe Storms Laboratory (NOAA NSSL) via internet. This product is a high-resolution tiled radar grid providing approximately 1-km resolution radar data every 5 minutes in a gridded netCDF file format. There are 8 files covering the spatial extent of the lower 48 states. These files provide composite reflectivity, echo top, velocity, precipitation, and vertically integrated liquid. The VDT 2.0 only ingests composite reflectivity. While this product continues to be useful (as an ancillary data source) in the initial stages of algorithm development, operational radar technology is becoming more sophisticated, and algorithm development for VDT 3.0 will leverage some of these new technologies to improve the remotely sensed information being included in product development. Specifically, we plan to explore the usefulness of new products from the upcoming dual-polarization upgrade.

The National Weather Service (NWS) has begun its initial upgrade of all Next Generation Radar (NEXRAD) operational radars to dual polarization. Conventional radar produces horizontally polarized pulses of microwave radiation to detect precipitation. With dual-polarization radars, both horizontally and vertically polarized pulses are produced, and the readings of these two polarizations allow the inference of particle type, such as liquid, ice crystals, or hail (Scharfenberg et al. 2005). An example of particle or hydrometeor identification compared to the conventional view by horizontal reflectivity is shown in Figure 5-1. On the right-hand image, the reflectivities are classified into a precipitation type, which is easier for the user to immediately understand and use to differentiate regions of, for example, hail (red) from heavy rain (dark green). The VDT 3.0 will be modified to allow ingestion of these data as they become available, and the usefulness of the new data, particularly as it relates to hydrometeor identification, will be examined for inclusion into the road impact algorithms. See Section 8 for discussion of how dual-polarization radar data may be incorporated into the VDT 3.0 algorithms.

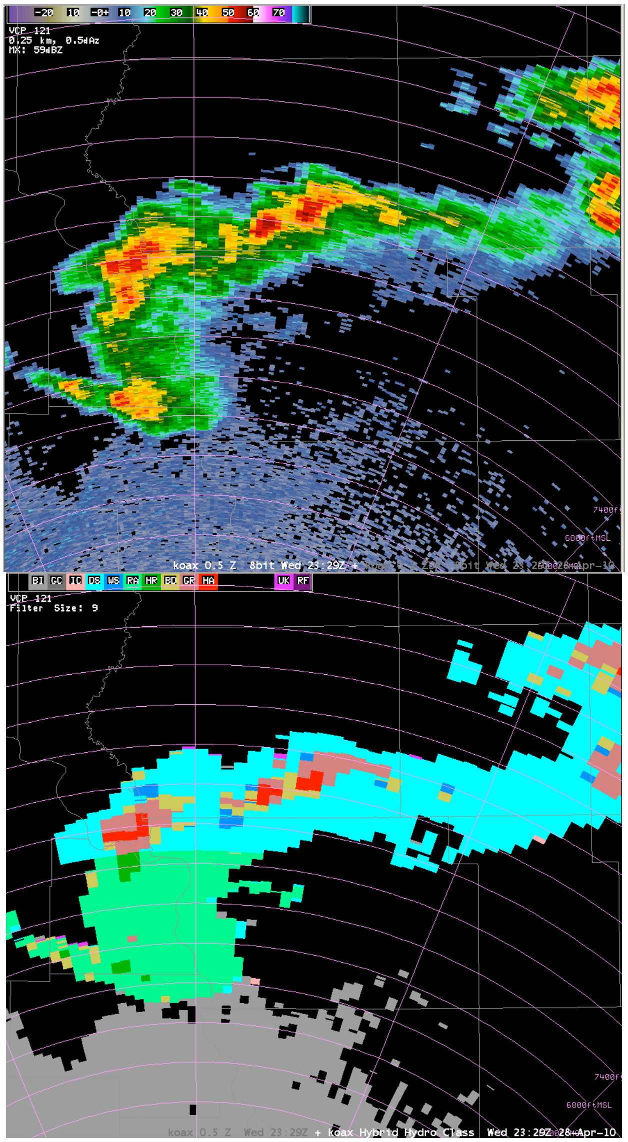


Figure 5-1**:** Example of horizontally polarized reflectivity (top) and hydrometeor classification from dual-polarization data (bottom). The top image is traditional reflectivity data, which does not provide much information on what type of precipitation may be falling. On the bottom, the reflectivities are classified into a precipitation type, which is easier for the user to immediately understand and use to differentiate regions of, for example, hail (red) from heavy rain (dark green). *Image courtesy the National Weather Service (http://www.roc.noaa.gov/WSR88D/)*

5.2. Gridded Weather Observation Products

In VDT 2.0, assimilated and gridded weather observations came from the Rapid Update Cycle (RUC) Surface Assimilation System (RSAS; Miller et al. 2002) in order to compare to mobile observations. As an upgrade to this ancillary data source in VDT 3.0, the RUC Real-Time Mesoscale Analysis (RTMA; Pondeca et al. 2007) will be used. The RTMA, like the RSAS, is an hourly analysis product, but it is a more representative diagnosis of the current state of the atmosphere than RSAS (e.g., Pondeca and

Manikin 2009). This product provides precipitation, atmospheric pressure, ceiling, dewpoint temperature, air temperature, wind speed, and wind direction on a 5-km continental USA grid.

5.3. Satellite

In VDT 2.0, satellite data are ingested from NOAA to help discriminate cloudy / clear conditions. This dataset consists of raw images of infrared channel 2 and 4 from the Geostationary Operational Environmental Satellite (GOES) East and GOES West satellites. These are merged and compared against known water / land areas to produce a 4-km gridded product that consists of a simple present / not-present five-minute instantaneous visible cloud mask in netCDF format. As an improvement to this existing cloud mask, the Naval Research Laboratory (NRL) cloud classification algorithm (Bankert 1994) will be examined for inclusion into VDT 3.0. These data are already being received at NCAR, and the VDT engine will be adjusted as needed to ingest them. Instead of a simple present / not-present classification, the NRL cloud classifier identifies the type of cloud, such as cirrus, stratus, or cumulonimbus (Figure 5-2). There are two different sets of classifications for day and night, with the daytime set taking advantage of both visible and infrared channels. As with the current cloud mask, the NRL cloud classifier data is gridded and at a 4-km resolution. See Section 8 for discussion of how the NRL cloud classifier may be incorporated into the VDT algorithms.

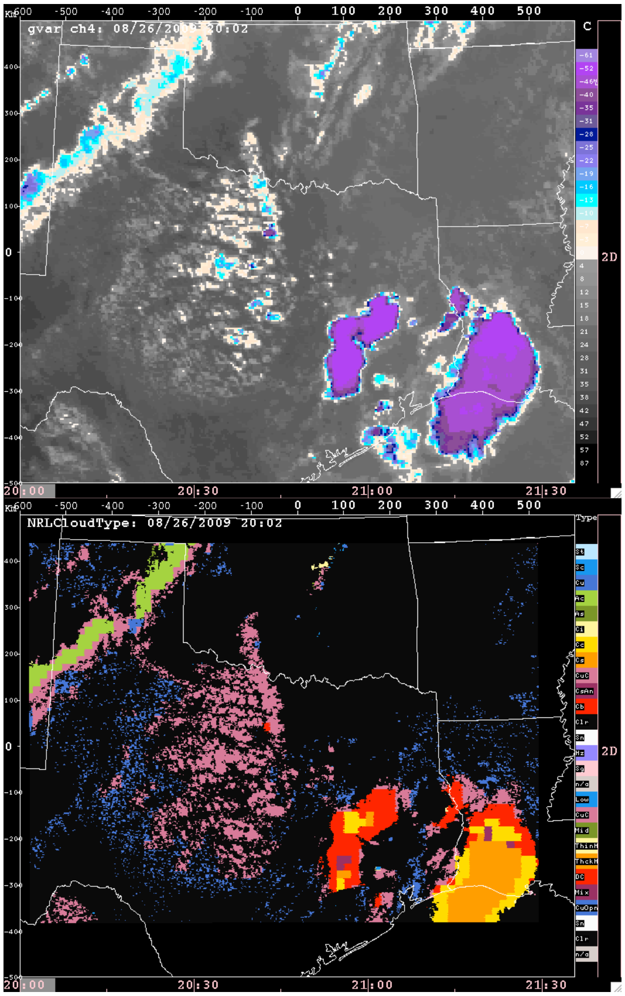


Figure 5-2: Comparison of channel 4 infrared satellite (top) and the NRL cloud classifier (bottom). The current cloud mask ingested into the VDT assigns present / not-present based on images such as the top. The NRL cloud classifier on the bottom assigns an inferred cloud type (cumulus in blue, altocumulus in light green, cumulus congestus in pink, cirrus in orange, and cumulonimbus in red). *Image courtesy Amanda Anderson.*

5.4. Surface Stations

In VDT 2.0, Automated Surface Observing System (ASOS) / Automated Weather Observing System (AWOS) data are received from the NWS Aviation Weather data feed. These data are organized into 5-minute netCDF files. As an upgrade to the fixed weather station data, in VDT 3.0, Road Weather Information System (RWIS) data from the *Clarus* network will be ingested, as well as a feed of non road-weather stations from the Meteorological Assimilation Data Ingest System (MADIS) and Weatherbug. As it becomes available, MoPED data will also be integrated into the VDT. This upgrade will enable more robust spatial QC of the mobile observations as well as provide a more dense observation field for inclusion in the various road hazard algorithms.

5.5. NWS Products

Storm watches and warnings issued by the NWS are updated at irregular intervals and issued through an Atom feed on the NWS website. This is an Extensible Markup Language (XML) file that is queried every 5 minutes and matched against the counties that the watch / warning is issued for. This data is temporarily placed into a PostgreSQL database for quick access. The archive format for these data is XML. For VDT 2.0, these warnings are simply included for use in a display module. For VDT 3.0, these warnings will be considered as input to the various road-hazard products. For example, a severe thunderstorm warning might indicate large hail is occurring, which could be added into the precipitation algorithm. At this time, the extent to which the algorithms will use these data is unknown. As the algorithm tuning process progresses, the usefulness of including these products as input will be examined.

Storm reports are updated infrequently from NOAA's Storm Prediction Center (SPC). The reports are made available via the web in Comma Separated Values (CSV) format after a storm has been reported. The reports are converted to a Javascript Object Notation (JSON) file for easy access, but the archive format is CSV. These data are not real-time but will be included in VDT 3.0 as information to the end-user for use in a display. This information could be useful in identifying areas to avoid when traveling, such as a section of road that may be covered in hail or a town that has sustained tornado damage.

Finally, each NWS office provides a daily forecast (Story of the Day) on its office website. These data are accessible via the web and the VDT 3.0 simply cross-references the Weather Forecast Office location with a link for the story, which can be shown on a display.

5.6. Social Media

Twitter weather reports are plain text reports provided by the public. These data are queried directly from Twitter, via the Twitter search Application Program Interface (API), every 5 minutes, and processed into a JSON file. These data can be automatically classified using a natural language processing (NLP) tool and integrated into the VDT. Moving beyond classification, these data can be analyzed for sentiment in order to indicate a negative or positive weather impact on a given area. The benefit of having these types of information in the short-term is strictly for displaying information to the end-user. In the long-term, as these types of information become increasingly available and trustworthy, they may be included as input into the automated road hazard products.

Chapter 6 VDT Quality Checking Routines

The QC routines are largely based on the *Clarus* QC routines, with modifications for mobile data. Only mobile data (CANBus and some add-on) are QC'd. This section outlines the specific QC tests in VDT 3.0. Like the *Clarus* System, all the tests are run on each sensor reading, and the test results are combined to obtain an overall confidence factor. This section describes the various QC tests.

6.1. *Clarus*-based Tests

6.1.1. Anticipated Range Test (ART)

The anticipated range test detects readings that fall outside the anticipated realistic range of sensor hardware specifications or theoretical limits (i.e., a maximum and minimum value). If the observation value is greater than or equal to the minimum, and less than or equal to the maximum, the sensor reading passes this quality check. If the sensor reading value is less than the minimum or greater than the maximum, the sensor reading does not pass. This test is useful in identifying observations that are likely not possible on the given sensor, particularly if the sensor uses an unusual value for identifying missing observations (e.g., 167.3 instead of -999). Unlike bounds for the *Clarus* System, it is not possible to know with precision the ART bounds for some vehicle instruments. This is because the OEMs do not release this information, so we cannot know with certainty the actual limits for every sensor on the road. Additionally, due to privacy / anonymity issues, even if we knew the ranges for sensors, we will not know which sensors are being used for some vehicles. The current valid bounds for version 2.0 of the VDT for the ART are[1]:

- Temperature: [-40,151°C]
- Barometric pressure: [580,1090 mb]
- Vehicle speed: [-734, 734 mph]
- Brake status: In bits, [0000, 1111] corresponding to all off, right rear active, right front active, left rear active, left front active, all on, or a combination of these.
- Brake boost: [0,2] corresponding to not equipped, off, and on.
- Wiper status: [0,5] and 255, corresponding to not equipped, off, intermittent, low, high, washer, and automatic present.
- Traction control: In bits, [00, 11] corresponding to not equipped, off, on, and engaged.
- Stability control: In bits, [00, 11] corresponding to not equipped, off, on, and engaged.
- ABS: In bits, [00, 11] corresponding to not equipped, off, on, and engaged.

---

[1] Some of these values seem unreasonably large, but they are the documented ranges. Additional investigation is warranted for VDT 3.0 and these limits will be updated in the future.

- Headlights: In bits, [0000-0000, 1111-1111] corresponding to parking lights on, fog lights on, daytime running lights on, automatic light control on, right turn signal on, left turn signal on, high beam headlights on, low beam headlights on, hazard signal on, all lights off, or a combination of these.
- Yaw rate: [0,655.35°/s]
- Horizontal acceleration latitudinal: [-44,44 mph]
- Horizontal acceleration longitudinal: [-44,44 mph]
- Steering wheel angle: [-655.36, 655.36°]
- Steering rate: [-381,381°/s]

These values will be reassessed as part of VDT 3.0 development, and additional categories added as new data are ingested during this project.

As part of revisions for version 3.0, we propose to output pass values as [1] and fail values as [0]. We will also explore using a gradation value from 0.0 to 1.0. This is a new test.

6.1.2. Time Step Test (TST)

The time step test detects sensor readings whose values change by more than a predefined variable-specific or station-specific rate over a thirty minute (past) and five minute (future) configurable period. For example, in the *Clarus* System, an air temperature reading from 2:00 p.m. will be compared to the corresponding air temperature sensor readings from the same sensor that was recorded in the time range of 1:30 p.m. to 2:05 p.m. This test requires that the sensor can be tracked through time, so this test can only work on vehicles that can be identified, such as fleet vehicles; it will not be useful to run this on passenger vehicle data that are anonymized.

Each time this test is invoked, it is given a single sensor reading. The system then obtains all of the sensor readings that have been received over the configured time period from the same sensor that are of the same weather parameter type. If either the current sensor reading or the prior sensor readings (a minimum of one is required) cannot be obtained, the test returns immediately with an error condition indicating that it was not able to run.

From the sensor, the system obtains configured positive and negative step threshold rates. If the difference between the current sensor value and the prior sensor value divided by the time difference in seconds ((current – prior) / time difference) falls between the negative step threshold and positive step threshold rates, then the current sensor reading passes the step quality check. If the computed rate falls outside the defined rates, then the current sensor reading does not pass the step quality check. This method assumes that the positive step threshold is specified as a positive value and the negative step threshold is specified as a negative value.

As part of revisions for version 3.0, we propose to output pass values as [1] and fail values as [0]. We will also explore using a gradation value from 0.0 to 1.0. This is a new test.

6.1.3. Persistence Test (PET)

The persistence test detects whether sensor readings remain constant for a predefined variable-specific period of time. For example, in the *Clarus* System, if consecutive pressure sensor readings remain unchanged to the precision of the instrument for four hours, the current sensor reading will not pass the persistence test. This test requires that the sensor can be tracked through time, so this test can only work on vehicles that can be identified, perhaps such as fleet vehicles; it will not be useful to run this on passenger vehicle data that are anonymized.

Each time this test is invoked, it is given a single sensor reading, which then determines the persistence time period. Consecutive sensor readings from the same sensor over that period of time preceding the current observation are then obtained. If the current sensor reading or the prior sensor reading cannot be obtained, the test returns immediately with an error condition indicating that it was unable to run.

If one or more of the consecutive sensor values are different, the current sensor reading passes the persistence quality check. If all of the consecutive sensor values over the given time period are equivalent, the current sensor does not pass the persistence quality check.

As part of revisions for version 3.0, we propose to output pass values as [1] and fail values as [0]. At present, we do not anticipate the PET as a good candidate for a non-binary output. This is a new test.

6.1.4. Spatial Tests – Air Pressure (STP) and Air and Pavement Temperature (STTa and STTp)

In version 2.0 of the VDT, both temperature and pressure were QC'd using the Nearest Station Test (NST). For the NST, a temperature observation passed if it was within 2°C of the station observation and pressure if the observation was within 5 mb. While this test was sufficient for the smaller DTE, a more sophisticated QC method is necessary for VDT 3.0.

In VDT 3.0, QC for air and pavement temperature and pressure will leverage work performed for the *Clarus* system specifically for spatial QC.

The Pressure Spatial Test (STP) will compare the observations with the closest surface stations in space and time. The nearest stations are currently defined as being within a 69-mile radius[2] and five-minute observation time from the data point; however, for the VDT 3.0, NCAR will examine other radius values and time ranges. In concept, this test mirrors the newest version of the Sea Level Pressure test (NCAR 2010b) performed in the *Clarus* System and is a major enhancement to the simple NST that VDT 2.0 uses for both temperature and pressure spatial QC.

The Temperature Spatial Tests (air and pavement; STTa and STTp) mirror the newer spatial QC from the *Clarus* system as well. Where there are more than 5 observations within a 69-mile radius and a five-minute observation time, the interquartile range (IQR) is used as a more robust method for spatial QC. In the instance where there are fewer observations than required for an IQR, the Barnes Spatial QC method will be used. These are new tests for the VDT 3.0.

Pass values are currently output as [1] and fails as [0]. As part of revisions for version 3.0, we propose to explore a less pass / fail or binary output. For example, we could envision values within 1°C of reference receiving a [1], values between 1°C and 2°C receiving a [0.5], and values outside 2°C receiving a [0]. Alternatively, the confidence could be altered by the test run as well, with the IQR

---

[2] The 69-mile radius is a legacy of earlier implementations of the *Clarus* QC algorithms.

receiving a higher confidence than the Barnes. The exact thresholds are also subject to change based on the collection of more data.

6.1.5. Climate Range Test (CRT)

The climate range test detects sensor readings that fall outside predetermined climate range values. This test does not exist in version 2.0 of the VDT, but it does for the *Clarus* system. We propose to mirror the *Clarus* CRT. Namely, the climate range data for *Clarus* were drawn from 30 years of National Centers for Environmental Prediction-Department of Energy (NCEP-DOE) Reanalysis 2 data. These reanalysis data are created by running a set of historical observational data through a common model, thus ensuring that the output data are consistent over time. The reanalysis also ensures that data are available in every time period at every grid point. Bounds for the CRT test were determined by computing monthly minimum and maximum values over a 2.5 degree x 2.5 degree fixed latitude-longitude grid. In the latitude band, this equates to a grid spacing of 172.5 miles. In the longitude band, this varies from 172.5 miles at the equator, to 0 miles at the poles. Each time this QC test is invoked, it is given a single sensor reading. The appropriate climate maximum and minimum values used for the test are determined by the month of the sensor reading date and the latitude / longitude region in which the sensor reading location falls. If the sensor reading value is greater than or equal to the climate minimum, and less than or equal to the climate maximum, the sensor reading passes the climate range quality check. If the sensor reading value is less than the climate minimum or greater than the climate maximum, the sensor reading does not pass the sensor climate range quality check. If it is determined that the monthly climate values are too coarse, then an additional enhancement to the VDT 3.0 version by using daily values will be considered.

As part of revisions for version 3.0, we propose to output pass values as [1] and fail values as [0]. At present, we do not anticipate the CRT as a good candidate for a non-binary output.

6.2. Non-*Clarus*-based Tests

6.2.1. Data Filtering Test (DFT)

In some cases, data may be obviously in error, such as latitude values greater than 90°N or a time value greater than 23:59:59. In special conditions, such as tunnels, data may be misleading (e.g., headlights and wiper status may differ in a tunnel). Additionally, some research (Mitretek 2006) indicates that at low speed, air temperature data may not be reliable, though other findings do not show this results (NCAR 2010b). This user-configurable QC will offer the ability to filter out these kinds of data.

6.2.2. Model Analysis Test (MAT)

The Model Analysis Test (MAT) compares the air temperature and pressure observations from the vehicles to those of a numerical weather model analysis field for the closest grid point. The VDT 2.0 is set up to use the RSAS data for the pixel closest in space to the observation. VDT 3.0 will use the more robust RTMA. These data will be ingested into the VDT 3.0 every hour. An analysis will be performed to determine the best technique for QC with this higher-resolution dataset. At this time, a nearest neighbor comparison is envisioned, but bilinear interpolation of multiple surrounding grid points will be examined as well.

Pass values are currently output as [1] and fails as [0]. As part of revisions for version 3.0, with FHWA agreement, we propose to explore a less pass/fail or binary output, such as in the PST, STTa, and STTp. There could be a time component here as well, with observations further in time from the RTMA time stamp receiving less confidence.

6.2.3. Neighboring Vehicle Test (NVT)

The Neighboring Vehicle Test (NVT) compares the given vehicle observation to neighboring vehicles in the road segment. Specifically, the standard deviation and the mean of the observations along a one-mile road segment during a five-minute snapshot are taken, and then each observation is checked to be sure it falls within a standard deviation, multiplied by a constant, of the mean of the road segment. Currently, the VDT version 2.0 uses a constant of 2.5, meaning the observations are checked to see if they fall within 2.5 standard deviations of the mean of the user-defined or standard one-mile road segment. This value was chosen based on previous tests with the Data Use and Analysis Project (DUAP) and the 2009 DTE experiment (DTE09) and 2010 DTE experiment (DTE10) data, which were tested with the VDT and determined 2.0 standard deviations was too strict. For VDT 3.0, this test will mirror the spatial tests (Barnes or IQR) being performed for the *Clarus* system and also include using ±5 minutes as the time period. A minimum number of vehicles are required for either the IQR or Barnes spatial tests to run in order for the NVT to be statistically sound. This may prove to be a difficult test to run until a larger vehicle dataset becomes available.

Pass values are currently output as [1] and fails as [0]. As part of revisions for version 3.0, we propose to explore a less deterministic output, such as in the PST, STTa, STTp, and MAT.

6.2.4. Combined Algorithm Test (CAT)

The Combined Algorithm Test (CAT) is designed to take the results of all the previous QC tests and combine them in order to assign a confidence to the observation that is being QC'd. There are multiple ways this may be achieved. First, one way is to simply assign a confidence based on the number of QC tests that the observation passed. For example, if the observation passes only the SRT, it could be classified as minimal confidence. A high confidence observation, on the other hand, would need to pass most of the tests, or else there could be a subset of tests that must be passed in order to flag the observation as high confidence. In version 2.0, the confidence flag is very rudimentary. The addition of dynamic scoring opens the possibility of much more in-depth confidence flags as well. As part of version 3.0, we propose to provide a scale from [0.00], which would indicate no confidence, to [1.00], which would imply complete confidence. This range would also be useful for ingesting data into numerical model schemes, such as the Weather Research and Forecasting (WRF) Real-Time Four Dimensional Data Assimilation (RTFDDA) scheme, which are built to handle confidence flags that range from 0.00 to 1.00. The exact process for generating the confidence value will be part of the analysis of the new data, but we anticipate a fuzzy logic approach.

Chapter 7 Statistics Module

The VDT statistics module remains largely unchanged in version 3.0. On a user-configurable road segment and time step (default to one-mile and five-minute), the statistics module provides mathematical outputs for a variety of mobile data elements, as shown in Table 7-1.

Table 7-1. Output from the Statistics Module by Road Segment.

	Mean	Median	Standard Deviation	Interquartile Range	Count	Min	Max	Number of Vehicles Activated
External air temperature	X	X	X	X	X	X	X	
Air pressure	X	X	X	X	X	X	X	
Wiper status					X			X
Headlight status					X			X
Anti-lock braking status					X			X
Traction control					X			X
Stability control					X			X
Rate of change of steering	X	X	X	X	X	X	X	X
Vehicle velocity	X	X	X	X	X	X	X	X
Brake status					X			X
Brake boost					X			X
Rain (rain sensor)					X			X
Relative humidity	X	X	X	X	X	X	X	X
Pavement temperature	X	X	X	X	X	X	X	X
RPM	X	X	X	X	X	X	X	X
Engine torque	X	X	X	X	X	X	X	X
Accelerometer	X	X	X	X	X	X	X	X

Chapter 8 Inference Module (VDT algorithms)

The Inference Module forms the core component of Stage III. This module provides derived data on road segments. It ingests QC'd mobile data and ancillary data and combines the two through a variety of scientific routines to develop the derived products discussed below.

8.1. Precipitation

The VDT 3.0 precipitation algorithm will provide an assessment of the type and intensity (amount / hour) or accumulation rate of precipitation that is falling to the road surface by road segment. We anticipate that it will have four precipitation types: rain, snow, ice/mixed, and hail. Additionally, it will distinguish between light/moderate and heavy rates of these precipitation types. "No precipitation" will, of course, also be included.

The following minimum input vehicle data are designed for inclusion into the algorithm:

- Air temperature
- Front wiper status
- Headlight status
- Ratio of vehicle speed to road segment speed limit

Additionally, the following ancillary data will be considered:

- Time of day
- Date
- Radar reflectivity
- Satellite cloud mask
- NRL cloud classifier
- Dual-polarization radar observations

Air temperature is a useful parameter in determining the likely type of precipitation (liquid vs. frozen). It can be useful for precipitation rate in that at very cold temperatures, heavy precipitation is unlikely due to the reduced amount of moisture in the atmosphere. Wiper status would, of course, indicate when the drivers needed to remove liquid or ice from their windshields. Other parameters would be useful in determining whether the wipers are removing precipitation or road splash, or if precipitation is occurring but wipers are not being used (e.g., light snow). Headlight status can give an indication of precipitation by indicating reduced visibility, as well as the common driving concept or legal requirement of turning on one's headlights in precipitation. Low speed ratios can indicate heavy precipitation as well as snowy or icy conditions. This would be especially helpful near the freezing point of 0°C, where lower speed ratios would indicate ice / snow and higher would indicate rain. Non-recurring congestion is important to consider too, and using time of day and date would help indicate if slow speed ratios are occurring during typical non-congestion times. Radar reflectivity will be useful particularly for indicating if precipitation is occurring and how heavily it is falling. Additionally, a new radar hail estimation technique (Kolodziej et al. 2010) will be investigated to assess whether radar data can accurately quantify hail conditions. For the satellite cloud mask, a sky clear of clouds can help differentiate between road splash and actual precipitation. When the NRL cloud classifier is incorporated, the cloud types can help identify storm cells (cumulonimbus) that are likely producing heavy rain or hail. This could be especially useful in areas where Doppler radar coverage is sparse, such as some areas of the high terrain in the western United States. Further into the future, when the National Weather Service begins running dual-polarization radar operationally, radar fields aside from the reflectivity can be analyzed for their ability to infer precipitation type. This could be useful for identifying precipitation type around 0°C. Additionally, dual-polarization radar should greatly improve rainfall rate estimation from the radar, useful in alerting for heavy rain and hydroplaning risk. It is important to note that the radar detects precipitation type above the ground rather than at the surface, and the further away from the radar the higher above the ground the detection. However, other parameters such as air temperature and vehicle observations can be used in conjunction with the hydrometeor identification above the ground to infer what is happening at the surface.

A variety of techniques will be examined to enhance the version 3.0 algorithm. In its current VDT version 2.0 form, the algorithm is a decision tree. Other methods such as fuzzy logic will be explored and their usefulness evaluated. Additionally, data mining procedures will be used to identify other useful vehicle observations and adjust the specific thresholds used within the algorithm. These will include, but are not limited to, developing a decision tree from a dataset of observations and linear regression models. Random forest, another type of classification and regression tree (CART; Breiman 2001), is an additional option.

Verification for precipitation algorithm development is a difficult problem. Voice recordings from DTE09 and DTE10 were useful for previous iterations, but they were also prone to time and location errors and they are influenced by subjective human interpretation. For example, one driver in a group of three may report rainfall as “heavy” while the other two drivers may report it as “light.” There is also a lack of description in some cases, such as a driver reporting that it is raining but not indicating whether the rain is light, moderate, or heavy. In Minnesota, plow drivers have the ability to report precipitation conditions, but it is unknown if this will be useful or not. Thus, surface observations, such as ASOS stations, rain gauges, and the NCEP Stage IV analyses (Lin et al. 2005) will be used. At some future point, we hope to perform tests on a controlled section of road with known environmental conditions to have good, accurate verification of precipitation, particularly freezing rain and snow.

The final output will be assigned a confidence flag to help the user understand how much confidence to place on the output. This confidence flag will be related to the verification done during the research and development.

8.2. Pavement Condition

The pavement algorithm is being developed to derive the pavement condition on a segment of roadway from the vehicle observations. Pavement conditions being considered are: dry, wet, road splash, snow, icy / slick, and hydroplaning risk. Particular effort will be focused on the slickness component.

The following vehicle data are being considered for this algorithm, although other vehicle information may also be considered:

- Air temperature
- Road temperature
- Front wiper status
- Ratio of vehicle speed to road segment speed limit
- Anti-lock brakes / traction control / stability control engagement
- Latitudinal acceleration
- Longitudinal acceleration
- Yaw rate
- Steering angle
- Steering angle rate

Additionally, the following non-vehicle data will be considered:

- Radar reflectivity
- Satellite cloud mask
- Dual-polarization radar observations
- NRL cloud classifier

Analysis of current data suggests that air and road temperature can help indicate whether roads are wet, snow-covered, and / or slick. For example, an air temperature of -6°C would support the possibility of a road being snow-covered, whereas an air temperature of 6°C would not be supportive of keeping snow or ice on the road. Front wiper status can help indicate precipitation or road splash. A low speed ratio can indicate a hazard, such a slick roadways, whereas a speed ratio near 1 would indicate roads are not very slick. Along with slickness, anti-lock brakes, traction control, or stability control are more likely to be engaged in such an environment than on dry pavement. Previous research shows that increased variability (i.e., larger interquartile ranges) of latitudinal and longitudinal acceleration, yaw rate, and steering angle corresponded to engagement of the anti-lock brakes and traction and stability control systems. This likely corresponds to a driver trying to control a swerving car, and hence the trigger to engage the systems. The rate of change of the steering angle will also be examined. For outside data sources, the presence or absence of radar reflectivity and clouds as viewed by the satellite can help distinguish between precipitation and road splash. The reflectivity is also helpful for identifying whether roadways are more likely to be dry or wet / snow-covered, and this distinction could further be improved with hydrometeor identification from dual-polarization radar. Also, as mentioned in the previous subsection, the NRL cloud classification could be useful in areas outside of (or with limited) radar coverage as well as identifying areas likely to be experiencing heavy precipitation, which in turn could indicate roadways that are becoming snow covered or have a hydroplaning risk.

As with precipitation, several techniques will be explored to form the pavement algorithm. Again, decision trees and fuzzy logic will be analyzed. Additionally, work has begun in using fuzzy logic sets for the algorithm. For this method, each pavement type (e.g., slick and dry) has a set of fuzzy logic functions associated with it. There is one function for each input parameter (e.g., wiper status) in each set. The value of each function (0 to 1) is added up for all functions within the set. The “most likely” pavement type is then the type that has the highest sum. For example, suppose the slick pavement set has a sum of 7 while the dry pavement set has a sum of 2. Between the two sets, the pavement is determined to be slick. This method was chosen based on a similar method used to determine the most likely hydrometeor type using dual-polarization radar measurements. Data mining will also be used as described in subsection 8.1.

Verification of these data will rely on cameras to be installed on the Minnesota fleet, available road-side cameras in Nevada, and trained driver reports. There seem to be a sufficient number of RWIS stations in Minnesota. Some report precipitation and surface temperature, as well as having cameras. There are also some with surface sensors that report wet, dry, etc. The RWIS in Nevada are more rudimentary, with only a few showing ‘yes/no’ for rain, but no further resolution. Additional details on verification will follow as more is known.

The final output will be assigned a confidence flag to help the user understand how much confidence to place on the output. This confidence flag will be related to the verification done during the research and development.

8.3. Visibility

The visibility algorithm is being designed to provide additional information by road segment as to both a general decrease in visibility as well as more specific visibility issues. This approach is being considered because specific visibility hazards such as blowing snow may be able to be derived, but in other cases the vehicle reports may indicate low visibility without an indication of the precise cause. In addition to reporting visibility as normal or low, possible specific hazards include dense fog, heavy rain, blowing snow, and smoke.

The following minimum input vehicle data will be considered for inclusion into the algorithm:

- Air temperature
- Front wiper status
- Headlight status
- Ratio of vehicle speed to road segment speed limit

Additionally, the following non-vehicle data will be considered:

- NRL cloud classifier
- Wind speed (currently from the nearest ASOS)
- Relative humidity (currently from the nearest ASOS)
- Visibility from nearest ASOS
- Output from the Precipitation algorithm

Analysis has determined that air temperature is a useful predictor for low visibility. This may be due to its inherent relation to relative humidity and thus haze and fog. Wiper status is a good indication of heavy precipitation or road splash causing a reduction in visibility. Headlight status can help identify both low visibility and fog. For example, drivers are likely to have their headlights on in low visibility, may turn their fog lights on, and are less likely to have their high beam lights on. Speed ratio may be useful in that drivers are likely to drive slower, especially on interstates or other highways, when visibility is low. The NRL cloud classifier could aid in identifying low cloud (e.g., stratus) associated with reduced visibility (e.g., haze or fog). Wind speed may be useful in both identifying the potential for blowing snow as well as the likelihood of fog, with higher wind speeds making fog less likely due to the mixing of the air. In analyzing the data, relative humidity has emerged as a useful predictor, likely due to its relation to haze, fog, and precipitation. Two methods may be used in assigning the relative humidity: directly from the nearby fixed weather stations, or using the stations' dewpoint temperature as well as vehicle air temperature. Visibility from the nearest fixed stations, of course, will give some idea of visibility in the vicinity of the road segment, but there are many local variations that may limit its usability. Finally, output from the precipitation algorithm may be useful in assigning specific precipitation-related visibility obstructions such as heavy rain or blowing snow.

As with the other algorithms, decision trees and fuzzy logic are being considered for this algorithm. Additionally, data mining methods will be used as described in subsection 8.1. Data mining has already proven useful in identifying parameters for the algorithm such as air temperature and establishing the importance of relative humidity.

Verification in the research and development phase is again difficult. As with pavement condition, cameras may be used particularly if they have a view of several fixed objects of known distance from the camera but this technique would be time-consuming and subjective. Most fixed weather stations report visibility, and this will also be useful, but must be used cautiously, as it cannot be used as a verification source if it is included as a parameter in the algorithm itself.

The final output will be assigned a confidence flag to help the user understand how much confidence to place on the output. This confidence flag will be related to the verification done during the research and development.

8.4. All-Hazards

The All-Hazards algorithm is a method that attempts to summarize all the hazards identified by the above three algorithms over individual road segments. This can manifest as a list of all hazards on a road segment / grid box or a ranking of the most significant hazards from low to high impact. The algorithm may be adjusted in the future as user feedback is analyzed or other algorithms adjusted.

Chapter 9 Open Source Philosophy

The VDT software suite emphasizes the use of open source software and avoids proprietary software packages unless such packages provide a significant advantage in initial development time or in establishing proof of concept. As development of the VDT software suite matures, it will be important to revisit all proprietary software components and determine if such components need to be replaced with open source alternatives.

There is a significant distinction between open source licenses. For example, the GNU General Public License (GPL) requires that all software developed using GPL libraries be made open source if the developed software is distributed. The GNU lesser GPL (LGPL) does not make such a requirement. The VDT currently emphasizes using LGPL libraries as opposed to GPL libraries. Thus, the Department of Transportation can make a determination whether the VDT software itself should be made open source.

The VDT is being developed on and operates using the open source Linux operating system. Linux has been available since the early 1990's and is widely used at universities and scientific research centers around the world. The underlying VDT code is written in C++, Java, JavaScript and Python. C++ is generally used for VDT data conversion, data ingest, probe data parsing, probe data sorting, probe data QC, the statistics module, the inference module, and in VDT output product formulation. The Java, JavaScript and Python languages are used in supporting web display. Python is also used as a glue layer language in running a sequence of VDT processing modules.

The VDT uses the open source GNU compiler collection (GCC) for compiling C++ code. This collection consists of front ends for C, C++, Objective-C, Fortran, Java, Ada and Go. Even though GCC provides support for Java, the VDT currently utilizes the free Java compiler from Oracle (www.java.com) for Java compilation.

The VDT also utilizes a number of third party libraries and applications, listed below.

Open Source Libraries Used by the VDT:

ann - approximate nearest neighbor
blitz - matrix library
boost - general C++ library
libconfig - configuration file library
extjs – JavaScript extension library. GPL license with commercial purchase option for non-open source software distributions
gdal - geospatial abstraction library. This library provides methods for transforming latitude and longitude data between different projections. Since the different input data sets arrive in different projections, GDAL is used unify the data using a common projection before being presented to the user.
geos – a library required by GDAL and Postgis for parsing spatial formats like WKT and WKB (http://en.wikipedia.org/wiki/Well-known_text)
netcdf - network common data format library. The VDT data files are stored using netCDF format.
proj4 - projection library required by GDAL and Postgis for transforming between spatial projections.

Python PIL - an image processing library that is used to build the images that are sent to the display.
udunits - units conversion library
xerces-c - XML parser
xmlbeansxx - XML library

Open Source Third Party Applications Used by the VDT:

PostgreSQL – a database server that is used by the VDT to store data ready to be sent to the display. The weather data stored in PostgreSQL will be transformed by the geo-spatial modules into web images that are presented to the user.

Postgis - an extension to Postgres that allows for making geo-spatial queries and transformations of data in the Postgres database. This is necessary for efficiently selecting and simplifying subsets of large datasets based on areas the user is currently interested in viewing.

Tomcat, Django and Apache – used for providing web services to the VDT display. Web services, in this instance, can be defined as a web location that takes input, such as location and time, and provides output in the form of an image or text that will be published directly on the VDT display and then presented to the user.

References

Bankert, R. L., 1994: Cloud classification of AVHRR imagery in maritime regions using a probablistic neural network. *J. Appl. Meteor.*, 33, 909–918.

Breiman, L., 2001: Random Forests. *Machine Learning*. 45, 5-32, doi: 10.1023 / A:1010933404324.

Kolodziej, A.G., V. Lakshmanan, T. Smith, and M. Richman, 2011: An automated technique to categorize storm type. *J. Atmos. Ocean. Tech*, submitted.

Lin, Y. and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications. Preprints, 19th Conf. on Hydrology, American Meteorological Society, San Diego, CA, 9-13 January 2005, Paper 1.2.

Miller, P.A., M.F. Barth, and L.L. Morone, 2002: NCEP RSAS Technical Procedures Bulletin, March 2002.

Mitretek, 2006: Vehicles as Mobile Sensing Platforms for Meteorological Observations: Introductory Research during a Winter Season. 267 pp.

NCAR, 2009a: An Analysis of Temperature and Pressure Data from Connected Vehicles in the Developmental Testbed Environment. FHWA-JPO-10-062/NTL #14968. 79 pp.

NCAR, 2009b: The Maintenance Decision Support System (MDSS) Prototype Release-6 Technical Description, Version 1.0. 167 pp.

NCAR, 2010a: *Clarus* Quality Checking Algorithm Documentation Report. 23 pp.

NCAR, 2010b: Vehicle-based Data Assessment. 61 pp.

Pondeca, M.S.F.V., G. S. Manikin, S. Y. Park, D. F. Parrish, W. S. Wu, G. DiMego, J. C. Derber, S. Benjamin, J. D. Horel, S. M. Lazarus, L. Anderson, B. Colman, G. E. Mann, and G. Mandt, 2007: The development of the real time mesoscale analysis system at NCEP. *23rd Conf. on IIPS*, San Antonia, TX, Amer. Meteor. Soc., P1.10.

Pondeca, M., and G.S. Manikin, 2009: Recent improvements to the Real-Time Mesoscale Analysis. Proceedings of the 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction. 1-5 June 2009, Omaha, NE.

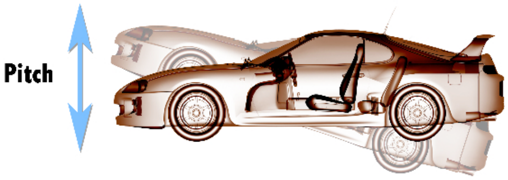
Scharfenberg, Kevin A., and Coauthors, 2005: The Joint Polarization Experiment: Polarimetric Radar in Forecasting and Warning Decision Making. *Wea. Forecasting*, 20, 775–788.

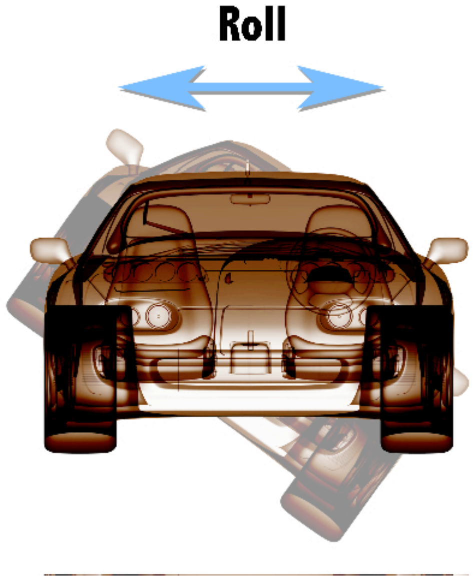
APPENDIX A. List of Acronyms

ABS	Antilock Braking System
API	Application Program Interface
ART	Anticipated Range Test
ASOS	Automated Surface Observing System
AWOS	Automated Weather Observing System
CANBus	Controller Access Network Bus
CAT	Combined Algorithm Test
CRT	Climate Range Test
CSV	Comma Separated Values
DFT	Data Filtering Test
DOE	Department of Energy
DOT	Department of Transportation
DTE	Development Test Environment
DTE09	2009 DTE Experiment
DTE10	2010 DTE Experiment
DUAP	Data Use Analysis and Processing
FHWA	Federal Highway Administration
GOES	Geostationary Operational Environmental Satellite
GPL	General Public License
IQR	Interquartile Range
JSON	JavaScript Object Notation
MADIS	Meteorological Assimilation Data Ingest System
MAT	Model Analysis Test
Mn/DOT	Minnesota DOT
MoPED	Mobile Platform Environmental Data
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NDOT	Nevada DOT
netCDF	Network Common Data Format
NEXRAD	Next Generation Radar
NLP	Natural Language Processing
NOAA	National Oceanic and Atmospheric Administration
NRL	Naval Research Laboratory
NSSL	National Severe Storms Laboratory
NST	Neighboring Station Test
NVT	Neighboring Vehicle Test
NWS	National Weather Service
OEM	Original Equipment Manufacturer
PET	Persistence Test
QC	Quality Check
RITA	Research and Innovative Technology Administration
RSAS	RUC Surface Assimilation Systems
RTFDDA	Real-Time Four Dimensional Data Assimilation
RTMA	Real-Time Mesoscale Analysis
RUC	Rapid Update Cycle
RWFS	Road Weather Forecast System
RWIS	Road Weather Information System

RWMP	Road Weather Management Program
SHRP2	Strategic Highway Research Program 2
SPC	Storm Prediction Center
STTa	Spatial Test – Air Temperature
STTp	Spatial Test – Pavement Temperature
STP	Spatial Test – Air Pressure
TST	Time Step Test
TOPR	Task Order Proposal Request
UCAR	University Corporation for Atmospheric Research
USDOT	U.S. Department of Transportation
VDT	Vehicle Data Translator
VIDAS	Vehicle-based Information and Data Acquisition System
WRF	Weather Research and Forecasting
XML	Extensible Markup Language

APPENDIX B. Glossary

Variable	Description
Accelerometer	A device that measures the acceleration of the vehicle in any direction.
Adjacent Snow Depth	The depth of snow on representative areas other than the highway pavement, avoiding drifts and plowed areas.
Anti-lock brakes	A system that prevents wheels from locking up under constant braking pressure
Atmospheric Pressure	The force per unit area exerted by the atmosphere.
Brake boost	A vacuum-assisted brake system that greatly increases the force applied to the vehicle's brakes.
Detected Friction	The measured coefficient of friction on the road surface.
Dewpoint Temperature	The temperature to which a given parcel of air must be cooled, at constant atmospheric pressure, for water vapor to condense into water.
Pitch	 *image courtesy Robert Marfuta and* *http://www.istockphoto.com/file_closeup.php?id=3466723*
Precipitation Indicator	Indicates whether or not moisture is detected by a sensor. “Precip” equals moisture is currently being detected; “noPrecip” equals moisture is not currently being detected; “error” means the sensor is either not connected, not reporting, or is indicating an error.
Precipitation Situation	Describes the weather situation in terms of precipitation. Intensity meaning: • slight < 2mm/hr water equivalent • moderate >= 2 and < 8 mm/hr water equivalent • heavy >= 8 mm/hr water equivalent
Rainfall or Water Equivalent of Snow	The rainfall, or water equivalent of snow, rate.
Rate of change of steering wheel	The angle of change in the steering wheel. Numbers greater than 360 indicate more than one revolution of the steering wheel.
Roadway Ice Thickness	Indicates the thickness of the ice on the driving surface.
Roadway Snow Depth	The current depth of unpacked snow on the driving surface.
Roadway Water Level Depth	Indicates the depth of the water on the driving surface.

Roll	 *image courtesy Robert Marfuta and http://www.istockphoto.com/file_closeup.php?id=3466723*
Solar Radiation	The ultraviolet, visible, and near-infrared (wavelength of less than 3.0 micrometers) radiation hitting the earth's surface.
Spot Wind Direction	The direction from which the wind is blowing at any instantaneous time.
Spot Wind Speed	The wind speed at any instantaneous time.
Stability control	A system that measure loss of stability in the vehicle and redirects power to certain wheels in an attempt to regain stability.
Surface Temperature	The current temperature on the driving surface.
Total Radiation	The average total radiation hitting the earth's surface.
Traction control	A system that helps to prevent loss of traction between the tires and the driving surface. May be related to ABS.

Yaw	 *image courtesy Robert Marfuta and http://www.istockphoto.com/file_closeup.php?id=3466723*

APPENDIX C. J2735 Information

We recommend the following variables be included as a “WeatherReport” dataframe in the next J2735 revision. Each of these variables should be assigned a unique parameter ID (PID). Most of these data elements are ported from the NTCIP 1204 standards, with some modifications for the mobile environment.

Data Frames:

Pressure/Wind

- Atmospheric Pressure
- Spot Wind Direction
- Spot Wind Speed

Temperature

- Air Temperature
- Dewpoint Temperature
- Surface Temperature

Radiation

- Solar Radiation
- Total Radiation

Precipitation

- Precipitation Indicator
- Rainfall or Water Equivalent of Snow
- Precipitation Situation
- Roadway Water Level Depth
- Adjacent Snow Depth
- Roadway Snow Depth
- Roadway Ice Thickness

Visibility

- Visibility

Friction

- Detected Friction

Variable	Description	Valid Range	Data Resolution	Temporal Resolution
Atmospheric Pressure	The force per unit area exerted by the atmosphere in 1/10ths of millibars, a.k.a. tenths of hectoPascals. A value of 65535 shall indicate an error condition or missing value.	650.0 mb – 1200.0 mb	INTEGER (0..65535)	Once every 20 seconds
Spot Wind Direction	The direction from which the wind is blowing measured in degrees clockwise from true North. A value of 361 shall indicate an error condition or missing value. The wind direction shall be corrected for vehicle movement.	0° - 359°	INTEGER (0..361)	Once every second
Spot Wind Speed	The wind speed in tenths of meters per second. The value of 65535 shall indicate an error condition or missing value. The wind speed shall be corrected for vehicle movement.	0.0 m/s – 250.0 m/s	INTEGER (0..65535)	Once every second
Air Temperature	The air temperature in tenths of degrees Celsius. The value 1001 shall indicate an error condition or missing value.	-100.0°C – 100.0°C	INTEGER (-1000..1001)	Once every 20 seconds
Dewpoint Temperature	The dewpoint temperature in tenths of degrees Celsius. The value 1001 shall indicate an error condition or missing value.	-100.0°C – 100.0°C	INTEGER (-1000..1001)	Once every 20 seconds
Solar Radiation	The ultraviolet, visible, and near-infrared (wavelength of less than 3.0 micrometers) radiation hitting the earth's surface in watts per square meter. The value of 701 shall indicate a missing value.	0 W/m^2 – 700 W/m^2	INTEGER (0,701)	Once every 20 seconds
Total Radiation	The average total radiation hitting the earth's surface in watts per square meter. The value of 1001 shall indicate a missing value.	0 W/m^2 – 1000 W/m^2	INTEGER (0,1001)	Once every 20 seconds
Visibility	Surface visibility measured in tenths of a meter. The value 200001 shall indicate an error condition or missing value.	0.0 m – 20000.0m	INTEGER (0..200001)	Once every 20 seconds

Surface Temperature	The current pavement surface temperature in tenths of degrees Celsius. The value 2001 shall indicate an error condition or missing value.	-100.0°C – 200.0°C	INTEGER (-1000..2001)	Once every second
Precipitation Indicator	Indicates whether or not moisture is detected by the sensor. “Precip” equals moisture is currently being detected; “noPrecip” equals moisture is not currently being detected; “error” means the sensor is either not connected, not reporting, or is indicating an error.	N/A	INTEGER { precip (1), noPrecip (2), error (3)}	Once every second
Rainfall or Water Equivalent of Snow	The rainfall, or water equivalent of snow, rate in tenths of grams per square meter per second. The value of 65535 shall indicate an error condition or missing value.	0.0 – 11.0 g/m^2/s	INTEGER (0..65535)	Once every second
Precipitation Situation	Describes the weather situation in terms of precipitation. Intensity meaning: • slight < 2mm/h water equivalent • moderate >= 2 and < 8 mm/h water equivalent • heavy >= 8 mm/h water equivalent	N/A	INTEGER { other (1), unknown (2), noPrecipitation (3), unidentifiedSlight (4), unidentifiedModerate (5), unidentifiedHeavy (6), snowSlight (7), snowModerate (8), snowHeavy (9), rainSlight (10), rainModerate (11), rainHeavy (12), frozenPrecipitationSlight (13), frozenPrecipitationModerate (14), frozenPrecipitationHeavy (15)}	Once every second
Detected Friction	Indicates measured coefficient of friction in percent. The value 101 shall indicate an error condition or missing value.	0 – 100	INTEGER (0..101)	Once every second
Roadway Water	Indicates the depth of the water on the	0 cm – 255 cm	BYTE (0..256)	Once every second

Level Depth	roadway in centimeters. The value 256 indicates an error or missing value.			
Adjacent Snow Depth	The depth of snow in centimeters on representative areas other than the highway pavement, avoiding drifts and plowed areas. The value 256 indicates an error or missing value.	0 cm – 255 cm	BYTE (0..256)	Once every second
Roadway Snow Depth	The current depth of unpacked snow in centimeters on the driving surface	0 cm – 255 cm	BYTE (0..256)	Once every second
Roadway Ice Thickness	Indicates the thickness of the ice in millimeters. The value 256 shall indicate an error condition or missing value.	0 mm – 255 mm	BYTE (0..256)	Once every second

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free “Help Line” 866-367-7487
www.its.dot.gov

[FHWA-JPO-11-127]

U.S. Department of Transportation
Federal Highway Administration
Research and Innovative Technology Administration