INTELLIDRIVE[SM] ROAD WEATHER RESEARCH & DEVELOPMENT - THE VEHICLE DATA TRANSLATOR

Sheldon Drobot
William P. Mahoney III
Elena Schuler
Gerry Wiener
Michael Chapman
National Center for Atmospheric Research
3450 Mitchell Lane
Boulder, CO 80301
303-497-2705, drobot@ucar.edu

Paul A. Pisano
Pat Kennedy
USDOT/Federal Highway Administration
1200 New Jersey Avenue, SE
Washington, D.C. 20590
202-366-1301, paul.pisano@dot.gov

Benjamin B. McKeever
USDOT/Research and Innovative Technology Administration
1200 New Jersey Avenue, SE
Washington, DC, 20590
202-366-4876, ben.mckeever@dot.gov

Andrew D. Stern[1]
Noblis, Inc.
3150 Fairview Park Drive South
Falls Church, VA 22042
703-610-1754, astern@noblis.org

ABSTRACT

One of the goals of the Research and Innovative Technology Administration's IntelliDrive[SM] initiative is for the public and private organizations that collect, process, and generate weather products to utilize vehicle sensor data to improve weather and road condition products. It is likely that some users will not be able to contend with the complexities associated with vehicle data, such as data quality, representativeness, and format. A solution for addressing this issue is to utilize a Vehicle Data Translator (VDT) to preprocess weather-related vehicle data before they are distributed to data subscribers. This paper will describe the VDT and how vehicle data sets are being processed by the prototype VDT to generate derived weather and road condition information.

KEY WORDS

Road Weather
Probe Messages

[1] Now at NOAA/National Weather Service

IntelliDrive^{SM}

INTRODUCTION

The utilization of data from mobile platforms is not new in the weather community, but the utilization of data from vehicles poses significant technical challenges (1–3), particularly with respect to data quality. Nevertheless, vehicle-based probe data from initiatives such as IntelliDrive^{SM} will significantly increase the density of weather observations near the surface and will provide unique datasets for deriving and inferring road and atmospheric condition information.

The amount of data that would potentially flow through a vehicle-based data network such as IntelliDrive^{SM} could be immense. It is likely that many prospective users will not be capable of handling the vast quantities of data that will be generated. Applications must be implemented to facilitate the use of vehicle-based weather data, because without such a function, the feasibility of utilizing vehicle probe data will be lower and there will be substantially more risk in its use.

One possible solution, a Vehicle Data Translator (VDT), is being developed to extract data elements needed to derive weather and road condition information from vehicle probe data; filter the data to remove samples that are likely to be unrepresentative; quality check the data utilizing other local surface observations and ancillary datasets; generate statistical output for specific road segments; and disseminate the quality-checked and statistically processed data to subscribers. These may include other data processing and dissemination systems, such as the U.S. Department of Transportation's (USDOT) *Clarus* System.

Under contract with the USDOT, research and development are being conducted by the National Center for Atmospheric Research (NCAR) to design, develop, and test the VDT concept. Data from the Detroit Developmental Test Environment (DTE) Proof of Concept (POC) test, DTE 2009 Winter Demonstration, and Michigan Data Use Analysis and Processing (DUAP) Project are being used to evaluate and test the prototype VDT.

This paper outlines recent progress on development of a VDT.

VEHICLE PROBE DATA - WEATHER DATA PROCESSING

Vehicle data can be complex and they pose a significant analytical challenge, particularly when it comes to measuring or deriving weather and road condition data. For example, some obvious issues center on how to deal with the large data volume, timeliness of the data, data quality and representativeness, and data format(s). These issues are not dissimilar to those associated with other fixed meteorological datasets, but the complexities are compounded because end users will have little knowledge about the data source. In contrast, the National Weather Service (NWS), Federal Aviation Administration (FAA), and other traditional providers of weather data follow stringent standards for instrumentation accuracy, precision, and siting.

There is a well-founded belief in the meteorological community that "bad data are worse than no data". End users have demonstrated that they need to be very comfortable with data quality before they will utilize new datasets. What does this mean for vehicle weather and road condition data? It is clear that it will take a significant amount of research and outreach to demonstrate that vehicle data are of sufficient quality to be used for operational purposes. The uncertainties and complexity associated with raw vehicle data (i.e., unprocessed data from vehicles) will likely deter many end users from using the data. As noted above, it is likely that most end users will not be able to handle the immense volume of vehicle data, let alone deal with data quality questions. We anticipate that many, if not most, users of vehicle weather data will require *processed* data. In this context, *processed* data means vehicle data that are extracted from the network of vehicle probe data, quality checked, and disseminated to active data subscribers in near real time. It is also likely that, due to the volume of data, many users will prefer statistically-derived data representing specific geographical areas (such as road segments) or times.

THE NCAR VEHICLE DATA TRANSLATOR CONCEPT

In a fully functional IntelliDrive^{SM} environment, millions of vehicles will be acting as probes and will continuously send reports to the vehicle data network. The concept of the weather and road condition data processor for vehicle data has been discussed in both *Clarus* and Vehicle Infrastructure Integration (VII) initiative meetings (VII was the predecessor to IntelliDrive^{SM}). In our view, not only is the concept sound, but the need for such a function is critical.

A conceptual illustration of the primary processing components of the proposed VDT is shown in Figure 1. Data from IntelliDrive^{SM}-equipped vehicles (e.g., personal and fleet vehicles) is communicated to the Road Side Units (RSEs) when the vehicles are within range of the receivers. The RSEs are connected to the IntelliDrive^{SM} communications network where most of the data will flow. Individual processing components of the VDT are described in the following sections.

Data Parser

The proposed VDT will include a data parser function that will extract relevant weather and road condition vehicle probe fields from the vehicle data network. The data elements selected for extraction will be determined by research results and feedback from stakeholders in both the atmospheric and surface transportation communities. The data elements could be added or subtracted as needs vary. The data flowing out of the data parser is still considered raw as it has not been processed in any way.

Data Filtering Algorithms

Data filtering algorithms will then be applied to chosen data elements to remove data that are not likely to be representative of the true conditions. For example, outside air temperature measurements may not be representative of the true ambient conditions if the vehicle speed is less than 25 mph (4). Therefore, one test that could be applied to the vehicle data is to discard all outside air temperature data measurements when the vehicle speed is less than 25 mph.

Filters could also be applied to data collected at particular locations that are known to generate errors (e.g., data measurements from inside tunnels). The process of deciding when and how to filter data will require considerable research and will need to be done with great care, as one would not want to remove data that may have some value.

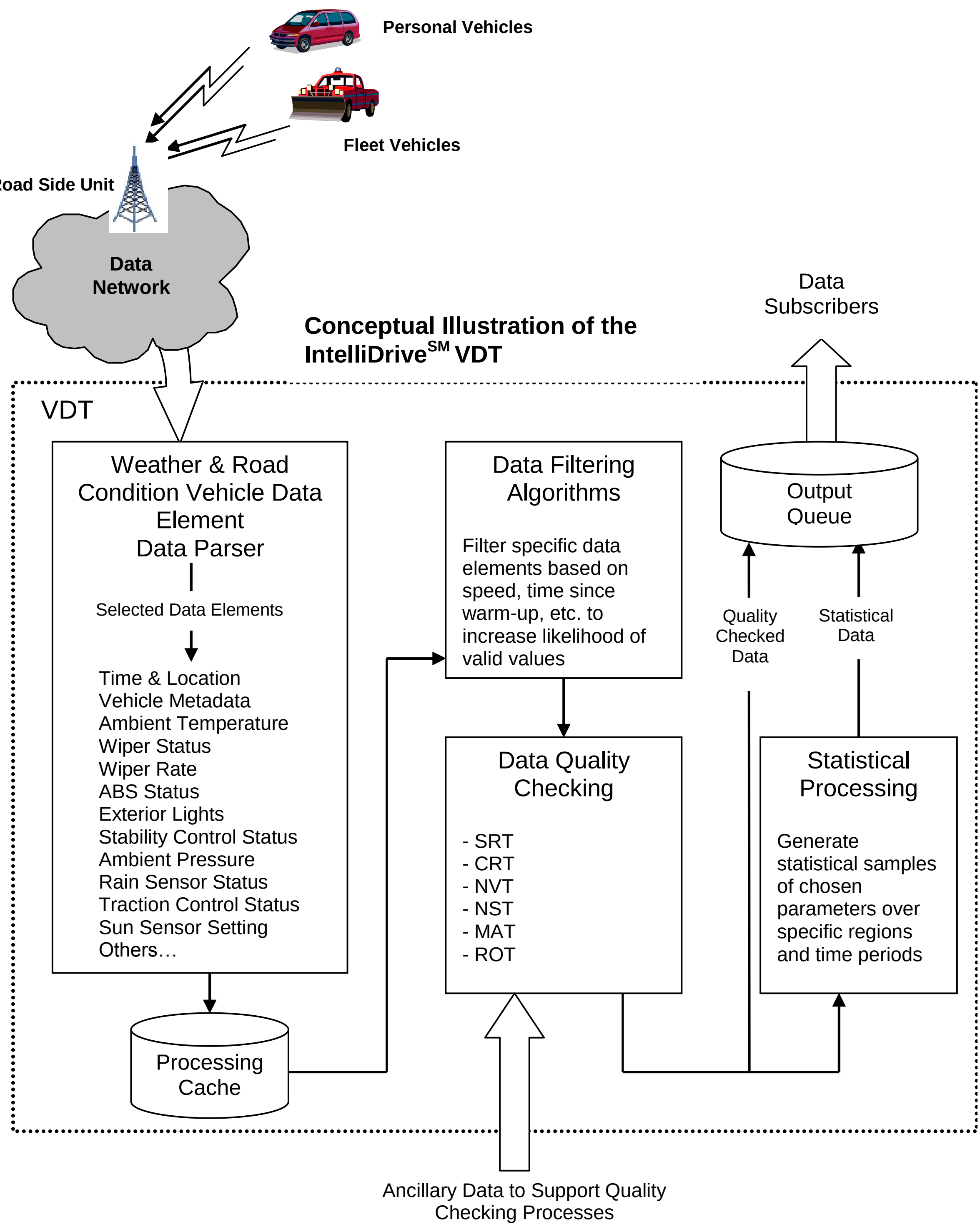

Figure 1. Vehicle Data Translator (VDT) Conceptual Illustration

<u>Data Quality Checking Algorithms</u>

A benefit of removing data that are considered unrepresentative in the filtering procedure is a reduction in the data volume that will need to undergo more complex and computationally intensive quality checking (QC) procedures. The QC tests will include many of the common tests that are applied to surface weather data and more complex tests to handle road condition data. In both cases, ancillary data, such as surface weather observations, satellite, radar, climatological data, and model output statistics will be required to conduct many of the quality tests. However, the VDT QC process will need to be sufficiently quick to ensure minimal latency.

The quality checking methods used in the VDT are related to some of those described in the *Clarus* system (5), and include the following:

- Sensor Range Test (SRT)
- Climatological Range Test (CRT)
- Neighboring Vehicle Test (NVT)
- Neighboring Surface Station Test (NST)
- Model Analysis Test (MAT)
- Remote Observation Test (ROT)

The SRT will be employed to identify observations that fall outside the range of the known sensor hardware specifications. Unlike the *Clarus* system, data processing algorithms included in the VDT will not have the advantage of “knowing” what type of sensor (e.g., make, model) produced each measurement. In order to develop an effective SRT, it will likely be necessary to conduct research on the automotive sensors that are available in the market place and make an educated assessment of the reasonable bounds for this test.

The CRT test will identify observations that fall outside of location-specific climatological (or historical) ranges. This is a more complex task than the SRT because of the variability of the climatological range values over times, dates, locations, and seasons. Once a climatological range is assigned for a specific road segment or geographic region (e.g., based on weekly values), the test will simply allow for observations within that range and flag observations that fall outside of the range.

The NVT is a “nearest neighbor” test which will compare the observation to neighboring vehicles in the road segment. If the observation value falls outside of a dynamic threshold, then the observation will fail the test. The threshold likely will be determined by the number of neighboring observations that are available for comparison. In other words, the higher the quantity of vehicles in a road segment, the tighter the threshold range.

The NST will compare neighboring surface station observations (e.g., Road Weather Information System (RWIS), Automated Surface Observing Systems (ASOS), Automated Weather Observing Systems (AWOS)) to each vehicle observation. This test will obviously depend on the availability of a surface observation within a suitable distance of the road segment.

The MAT will compare observations with a model surface analysis, such as the Real-Time Mesoscale Analysis (RTMA) or the Rapid Update Cycle (RUC) Surface Assimilation Systems (RSAS). This test will compare the observed value to either a range of grid values along a road segment or within a predetermined threshold of the grid value closest to the location of the vehicle.

The ROT will compare vehicle observations to remotely sensed data from either satellite and/or radar. Both the radar and satellite can be used as either a complimentary step with the other QC tests or as the final option if no neighboring vehicles or surface stations are available. As an example, pavement temperature, rain and sun sensor observations are good candidates for this test.

In the QC process, data quality flags will be applied to the raw vehicle data so that data subscribers will have the flexibility of utilizing the raw data or taking advantage of the quality checking flags. After the QC procedure, some data will flow directly to the output queue to minimize data latency (Figure 1).

Statistical Processing and Derived Variables

Some of the data will be cached and processed to generate statistical values for a given location (e.g., grid cell or point) and time period. The statistical processing will create two separate data streams: "processed data" and "derived data". To generate processed data the application of a statistical technique (e.g., mean, median) to a set of observations over a known grid segment is performed, and it should result in a more robust sample and reduce the overall data load for users that either cannot or do not want individual vehicle data. To generate derived data, a more computationally intensive procedure is necessary that focuses on deriving new or enhanced road and atmospheric variables of interest to the surface transportation community (6). The derived variables will be a blend of weather and non-weather related IntelliDrive^{SM} data elements in conjunction with ancillary data sets. Table 1 lists possible derived observations that are currently being considered during the incipient phases of IntelliDrive^{SM}, the ancillary data required to construct each derived observation, and the relevant vehicle observations (observed and input).

Table 1 – Sample Matrix for VDT Derived Weather-Related Observations

Derived Observation	Ancillary Data	Vehicle Observation
Ambient Air Temperature	RSAS, Surface Stations	Vehicle Speed, Temperature, Hours of Operation
Barometric Pressure	RSAS, Surface Stations	Vehicle Speed, Barometric pressure, Hours of Operation
Precipitation Occurrence	Radar, Surface Stations, Satellite	Wiper Status, Vehicle Speed, ABS, Stability Control, Traction Control, Headlight Status
Precipitation Type	Radar, Surface Stations, RSAS, Sounding	Wiper Status, Vehicle Speed, Temperature, Elevation, ABS, Stability Control, Traction Control
Precipitation Rate	Radar, Surface Stations	Wiper Status, Rain Sensor, Vehicle Speed, ABS, Stability Control, Traction Control, Headlight Status
Fog	Radar, Surface Station Visibility, Satellite Cloud Classification	Wiper Status, Rain Sensor, Vehicle Speed
Pavement Conditions (wet, dry, icy)	Radar, Surface Stations	Vehicle Speed, ABS, Stability Control, Traction Control, Temperature, Wiper Status
Boundary Layer Water Vapor	Radar, Soundings, Surface Stations	Temperature, Wiper Status
Pavement Temperature	Surface Stations	Temperature, Vehicle Speed
Smoke	Satellite Smoke Algorithm, Surface Station Visibility,	Vehicle Speed, Headlight Status

An evaluation of the optimum road segment length will be performed to determine the appropriate value that will be required to generate a spatially representative derived observation value. After the road segment size is determined, a step will be needed to evaluate whether the number of observations over the valid road segment is sufficient to derive an observation. For example, if a road segment value is produced by incorporating data from only one vehicle at a given time, it is safe to assume that the confidence in that value will be significantly lower than if the value for a given segment and time was derived from multiple observations from multiple sources. It will be important to identify the ideal minimum number of observations during the early stages of VDT development; this number will differ for each type of observation.

It should also be noted that the merging of different data sources (vehicle observations and ancillary data) will require the use of expert systems (e.g., fuzzy logic, neural networks) and/or decision tree algorithms in order to produce robust observations. It is likely that a combination of techniques will be required for some of the products.

VDT PROCESSING REQUIREMENTS

The shear volume of data flowing through vehicle-based data networks, such as the one envisioned by IntelliDrive^{SM}, will require a communications infrastructure with substantial bandwidth and computational capacity to ensure that latency will not become a barrier to its use. Like the Internet, the technical capabilities, number of users, and applications of the system will expand with time. The design of the VDT will have to take into account the evolution of vehicle-based data networks and be extensible to handle growing processing needs.

Although continuing increases in computing power will reduce data latency concerns, some initial estimates of the VDT computational requirements are provided. In the determination of the VDT computational requirements, the following simplifying assumptions are made to constrain the problem:

- Input and output streams are not compressed
- Input and output record size of 40 bytes and 50 bytes, respectively
- Data archive files on disk are not compressed
- Data flowing into the VDT are parsed to include only weather-related data elements
- Only weather-related data elements are processed
- Statistical processing is done on a 2 km X 2 km grid covering the contiguous U.S. One representative value for each grid cell in each weather-related data element is computed every 15 minutes
- VDT output includes both raw vehicle data with quality checking flags and statistically processed data for road segments
- Ancillary data used in statistical processing includes:
    - Surface observations
    - Radar data
    - Satellite data
    - Weekly climatology
    - Model output statistics
- Five basic quality checking tests are performed on each data element
- Data from 1 million vehicles are processed
    - Rural data rate – one record every 20 seconds from each vehicle

- Urban data rate – one record every 4 seconds from each vehicle

Given these assumptions, for periodic data coming from 1 million vehicles operating in a rural area, a computer node with two 3 GHz dual core processors and 4 MB of cache, 6 GB of memory, 600 GB of disk space, and two 100 Mb Ethernet cards would be required. The cost of this hardware would be about $6,000. Data and input into the VDT would average 2 MB per second, and quality checked output would be on the order of 2.5 MB per second. Gridded output would be approximately 351 MB per hour. It would require 432 GB disk space to keep a two-day archive of weather-related vehicle data, including quality checking flags; a two-day archive of gridded data would require 16.8 GB of disk space.

To process urban data from 1 million vehicles, the minimal hardware requirements would include a computer node complete with two 3 GHz dual core processors and 4 MB of cache, 16 GB of memory, 2.5 TB of disk space, and three to four 100 Mb Ethernet cards. The input data rate would average 10 MB per second, while quality checked output data would be roughly 12.5 MB per second. The amount of data generated for each road segment every hour would remain the same at 351 MB; however, it would require 11.25 GB of memory to handle the data processing needs compared to 2.25 GB for the rural case. The hardware costs for a system with these characteristics are estimated at $9,000.

Ancillary data used in the quality checking process would require a single node consisting of two 3 GHz dual core processors with 4MB of cache, 16 GB of memory, 1.5 TB storage capacity, and a 100 Mb ethernet card. Presently, the approximate price of the system would be about $8,000.

The estimated hardware specifics, data rates, and storage requirements are supplied here to give the reader a general sense of what processing capacity would be required to effectively process weather-related IntelliDrive^{SM}-enabled data using a VDT. These estimates are based on several assumptions. Although all of these assumptions have an impact on the estimates, two are very central in terms of the amount of data processed in the VDT. The first is that the data originates from 1 million vehicles in rural or urban environments. However, in the U.S. there are roughly 245 million registered vehicles (www.bts.gov), so as IntelliDrive^{SM}-related technologies are deployed and implemented, the data flowing through the IntelliDrive^{SM} network will originate from more than one million vehicles. The second is that the VDT input data will include only weather-related elements, which will not likely be the case. The VDT will have to parse raw records (snapshots) before doing any quality checking or statistical processing. This would mean that the size of the input data records would be significantly larger. Increases in data volume can be handled by adding additional nodes.

The amount of data produced by vehicles in a mature IntelliDrive^{SM} environment will be considerable. However, a gradual adoption of IntelliDrive^{SM} is anticipated (7). As a result, computer storage capacity, memory, and processing speeds likely will be able to keep pace with the growth and development of IntelliDrive^{SM}. The main factors that may constrain the use of IntelliDrive^{SM}-enabled data network are device input/output (I/O). Testing using simulated data from 10 million vehicles revealed the time needed for disk I/O (roughly 7 to 8 minutes) was 3 to 4 times that required for road segment data processing and quality control. As more and more vehicle data are generated, the capacity needed to acquire, distribute, read, and write data will increase.

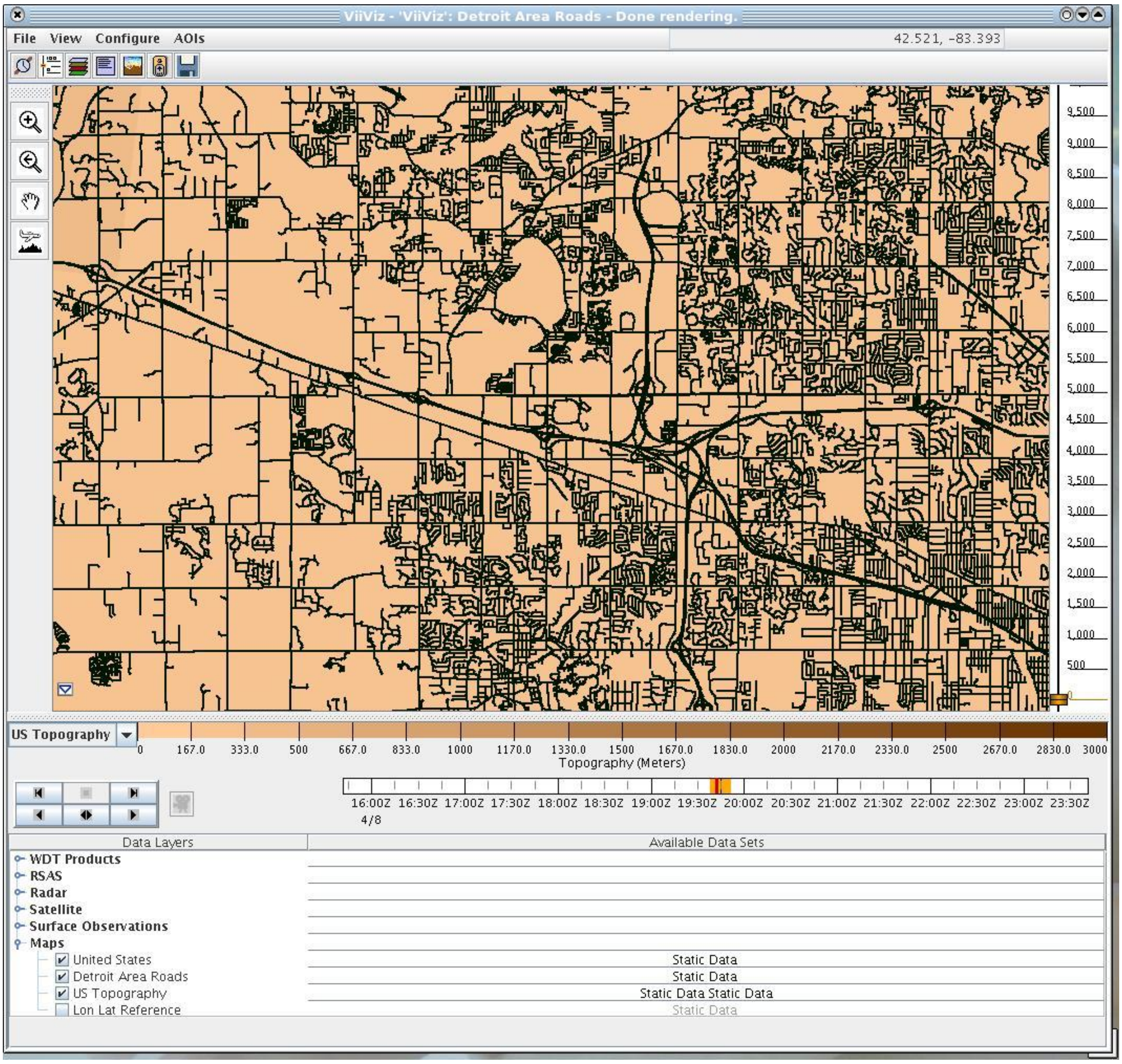

Figure 2. Example of prototype VDT display.

VDT DISPLAY

A key component of the VDT concept includes the capability to display the data. A prototype display is under development to support VDT research and to convey the utility of vehicle-based data. The VDT display is Java-based and it will be initially configured to display vehicle probe data sets covering the Detroit DTE region. The display can be re-configured to support other locations using an XML configuration file. The display will be capable of rendering the following data types:

- Vehicle probe messages
- Standard surface weather observations
- Gridded radar data
- RSAS data
- Satellite cloud mask data
- Road segment statistics (data range, variance, means, modes, etc.)

- Derived road segment values

The display will be divided into 4 major areas, a toolbar on top, a map area in the center, a time slider window below the map and a data layer section on the bottom (Figure 2). The data layer section will display the available datasets using a list with checkboxes. The user will be able to select a dataset by checking the corresponding data layer and selecting a time of interest in the time slider section. Point observations, such as METAR reports and probe messages, will be displayed as points on the display. The user will be able to click on a display point to bring up underlying data values in text form. In order to display road segment statistics and weather, the user can click on or mouse over the appropriate road segments.

VDT PRODUCT EXAMPLES

This section outlines two examples of VDT products. The first is a conceptual description of a derived fog product, and the second is a conceptual description of a precipitation product.

Fog Example

Fog remains one of the more difficult-to-diagnose road weather hazards. A schematic of a decision tree for the derivation of fog is presented in Figure 3. Probe data will provide useful observations of the status of the vehicle (e.g., wipers, vehicle speed) and observations of the atmosphere (e.g., temperature, rain sensor). These observations can be combined with ancillary data (e.g., radar, visibility and humidity from surface stations, and satellite cloud classification information) to diagnose the likelihood of fog. For example, if a vehicle's (and surrounding vehicles') speed decreases rapidly relative to the posted speed limit of a road segment, a decision tree can be used to determine if the decrease in speed was caused by fog, precipitation, or some other non-weather-related factor. The first test would determine the ratio of the rate of speed versus the speed limit of the road segment. If this ratio equals a certain threshold (e.g., 50%) then the algorithm would next consider the headlight status observation. If the headlights were on and the date and time suggest that it is daytime, then the algorithm would consider the wiper status. The algorithm would next check the relative humidity and visibility from the closest (in time) surface station observation report. If visibility is less than a predetermined threshold (e.g. ¼ mile) and humidity is high then there is some confidence that fog exists over the road segment. If a surface station is not available or if the observation from the surface station is outside of a predetermined temporal threshold then the algorithm could employ radar and satellite data. In this event, if the radar shows no precipitation but the cloud classification algorithm diagnoses a low-level stratus cloud, a fog classification could occur. While the above description assumes a decision tree for the combination of these data sources, a more robust fuzzy logic technique could also be employed when it is determined that applying discreet thresholds on the observations is not ideal.

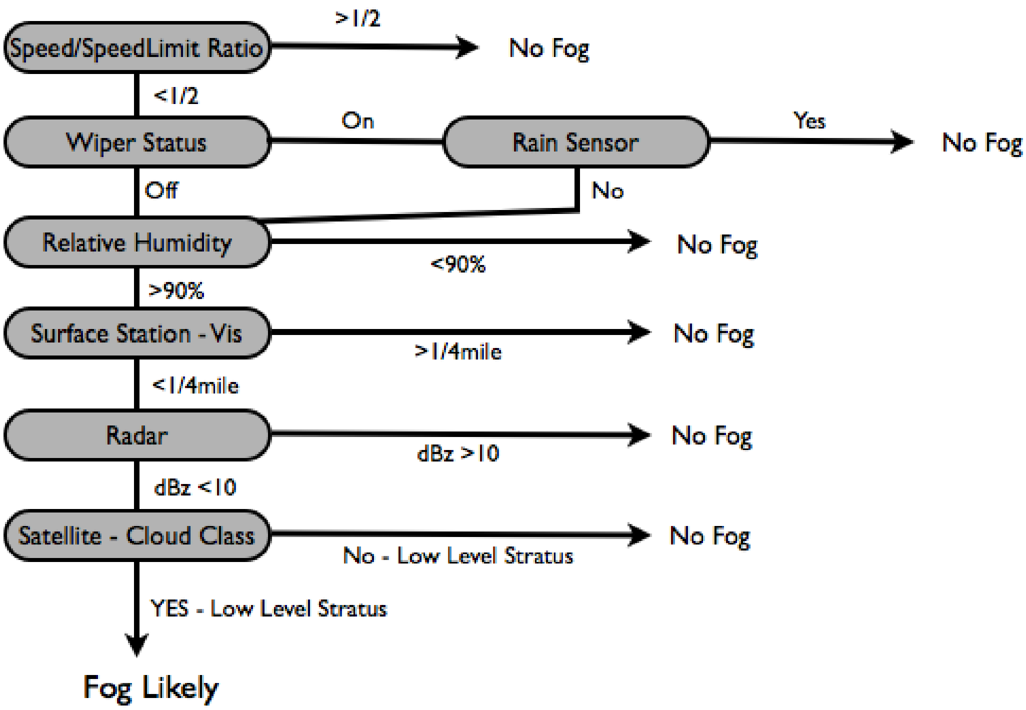

Figure 3. Schematic of decision tree algorithm for dense fog diagnostic application.

<u>Precipitation Example</u>

Another major hazard for the surface transportation industry is precipitation, due to the negative effects of precipitation on the tire/roadway interface (i.e., lower friction, hydroplaning). Diagnosing the occurrence, rate, and type of precipitation is a necessity if improvements to the safety and efficiency of the roadway are to be realized in the future. IntelliDrive^{SM}-enabled weather observations have the potential to provide the high density, real-time observations that could fill in the gaps between radar and the less dense fixed surface weather station network, and also add valuable information concerning the effects that the precipitation is having on the vehicle itself (e.g., ABS, activation of Traction Control, reduction in speed, collisions, etc.).

Figure 4 is a schematic, which represents the design for a precipitation algorithm using fuzzy logic. The algorithm would combine Wiper State with radar, surface observations, model data (e.g., RSAS), and satellite at different interest levels. Each variable would then be given a weight and fused into a consolidated interest field for precipitation. The final product would be a mapping of impacted road segments.

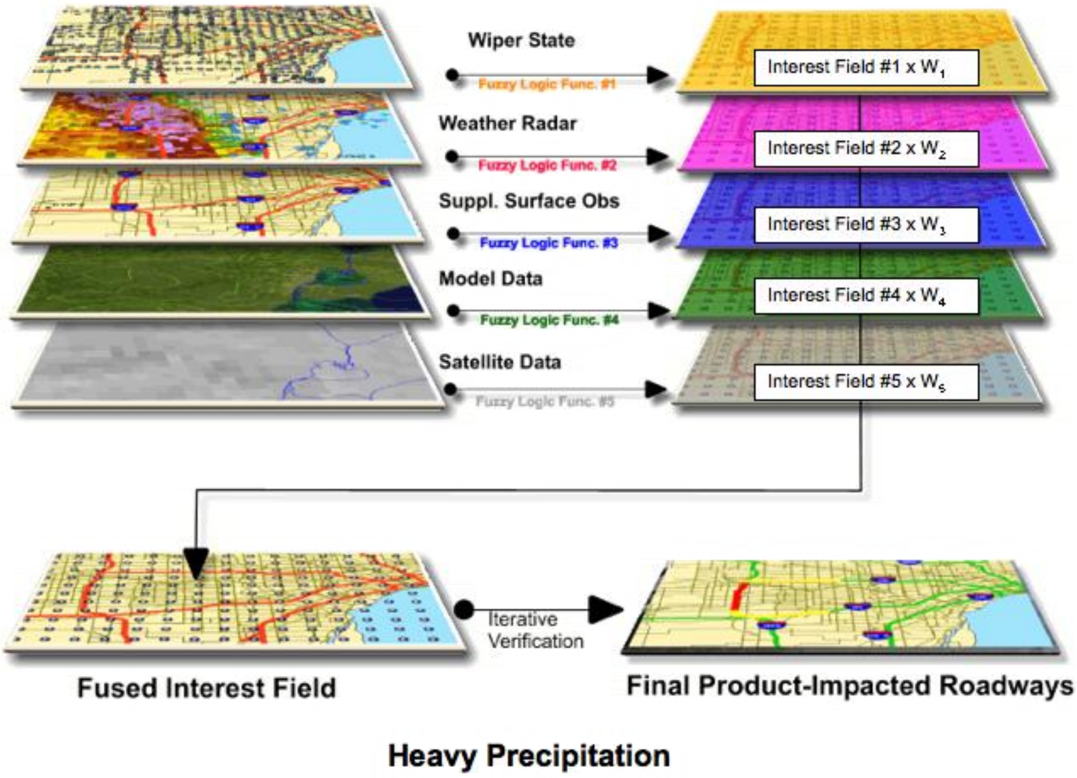

Figure 4. Schematic of the combination of vehicle data with ancillary data to produce a precipitation occurrence algorithm**.**

CONCLUSIONS

From a weather perspective, the overarching goal of the IntelliDrive^{SM} initiative is for the public and private organizations that collect, process, and generate weather and transportation products to utilize vehicle data to improve weather and road condition products, and then to provide those products to transportation system decision-makers, including travelers. Nonetheless, the utilization of data from vehicles poses significant technical challenges, particularly with respect to data quality and quantity. The amount of data potentially flowing through a vehicle-based data network, such as IntelliDrive^{SM}, could be immense and it is likely that many prospective users will not be capable of handling the vast quantities of data that are expected.

The VDT discussed in this paper is one approach to preprocess weather-related vehicle data before they are distributed to data subscribers. The function of the VDT is to extract data elements needed to derive weather and road condition information from vehicle probe data, filter the data to remove samples that are likely to be unrepresentative, quality check the data utilizing other local surface observations and ancillary datasets, generate statistical output for specific road segments, and disseminate the quality-checked and statistically processed data to subscribers, which may include other data processing and dissemination systems such as the U.S. Department of Transportation's (USDOT) *Clarus* System.

ACKNOWLEDGEMENTS

Research conducted during this project was funded by FHWA Contract DTFH61-08-C-00004.

REFERENCES

(1) K.K Srinivasan and P. P. Jovanis, “Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network”, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1537, Transportation Research Board of the National Academies, Washington, D.C., 1996, pp. 15–22.

(2) R.L. Cheu, H. Qi, and D.-H. Lee, “Mobile Sensor and Sample-Based Algorithm for Freeway Incident Detection”, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1811, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 12–20.

(3) S.E. Shladover and T.M. Kuhn, “Traffic Probe Data Processing for Full-Scale Deployment of Vehicle–Infrastructure Integration”, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2086, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 115–123. DOI: 10.3141/2086-14

(4) A.D. Stern, P. Pisano, P. Kennedy, K. Petty, W. Mahoney. “A next generation of observations based on passenger vehicles.” American Meteorological Society Annual Meeting, New Orleans, Louisiana, January 24, 2008, and published in the *88th Annual Meeting of the American Meteorological Society Proceedings*.

(5) United States Department of Transportation, “*Clarus* Weather System Design: System Design Description Version 1.1”, USDOT, ITS Joint Program Office. Washington, D.C. 2006.

(6) K.R. Petty and W.P Mahoney, “Weather Applications and Products Enabled Through Vehicle Infrastructure Integration (VII): Feasibility and Concept Development Study”, National Center for Atmospheric Research, Boulder, Colorado. 2007. 116 pp.

(7) T. McGukin. “Small town Attitude”, *ITS International NA Edition*, November/December 2007, pp. NA5-NA6.