2A.3 QUALITY OF MOBILE OBSERVATIONS COLLECTED DURING THE 2010 DEVELOPMENT TESTBED ENVIRONMENT EXPERIMENT

Amanda R. S. Anderson*, Michael Chapman, Sheldon D. Drobot, Alemu Tadesse
National Center for Atmospheric Research, Boulder, Colorado

Paul Pisano
Federal Highway Administration, Washington, District of Columbia

1. INTRODUCTION

On average, there are approximately 7100 weather-related vehicle fatalities in the United States each year (Noblis). By contrast, there are an average 574 deaths per year resulting from heat, floods, tornadoes, wind, lightning, winter weather, cold, and hurricanes combined (National Weather Service 2010). The number of vehicle fatalities involving adverse weather necessitates work on analyzing weather conditions along roadways to provide travelers with decision-making support to reduce the likelihood of a crash. The use of vehicle sensor data could revolutionize the delivery of road weather information to transportation decision-makers, including travelers.

To assess the quality of available vehicle sensor data, the 2010 Development Testbed Environment Experiment (DTE10) was run to provide an examinable dataset. This experiment ran over 19 days spanning 28 January to 29 March 2010. On each testing day, vehicles were driven on predetermined routes in the DTE area of Novi, Michigan (a suburb of Detroit). Testing days were chosen to encompass a variety of weather and driving conditions, including cold temperatures, clear conditions, heavy snow, rain, congestion, and rural routes. More detailed information about DTE10 is provided in Section 2.

This paper is divided into 6 sections. Section 2 describes the DTE10 experiment and the data collected from the vehicles. Section 3 presents the results of quality checking, including sensitivity tests, and statistics involving comparison between the mobile observations and a nearby weather station. Section 4 provides a summary and offers conclusions based on the results. Section 5 lists acknowledgements, and Section 6 lists references for this paper.

2. DATA

The DTE10 dataset contains over 239,000 air temperature and barometric pressure observations collected over 19 testing days from 9 vehicles. There were 3 Ford Edges and 6 Jeep Grand Cherokees. The full set of data collected from the vehicles is listed in Table 1.

In addition to the vehicle observations, air temperature, dewpoint temperature, and pavement temperature data were collected by Vaisala Surface Patrol HD units (for more information see Vaisala 2010a), which were mounted to the left-front quarter panel of each vehicle. These units added the value of providing dewpoint and pavement temperature, not currently measured by the vehicles, collocated with vehicle observations. A fixed weather station, the Vaisala WXT520 (for more information see Vaisala 2010b), was set up at the test facility to capture more representative surface weather data than the Detroit (KDTW) Automated Surface Observing System (ASOS) station, located about 30 miles away from testing. Initially, it was thought to use the Surface Patrol HD sensors as a primary source of validation and the WXT520 as a secondary source, but time series of the HD units showed some unrealistic fluctuations in air temperature. Additionally, the Surface Patrol HD units tended to become dirt, slush, and ice covered under various conditions. These issues, coupled with the lack of barometric pressure measurements, resulted in choosing the WXT520 observations as the ground truth measurements for this study.

Vehicle sensor observations and Surface Patrol HD measurements were collected from the vehicles via the On-board Equipment (OBE), with part of these logs being in Extensible Markup Language (XML) format. The logs were parsed into comma-delimited files for easier reading. The WXT520 observations were recorded via a single-board computer that was hooked up to the sensor.

Quality Checking (QCh) algorithms were applied to the vehicle sensor observations. These algorithms are integrated into the Vehicle

---

*Corresponding author address: Amanda R. S. Anderson, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307;
email: aander@ucar.edu

Data Translator (VDT, Drobot et al. 2009), which ingests vehicle observations and processes them with ancillary data such as radar, satellite, and ASOS observations and assigns them to a specific 1-mile segment of road network per 5-minute interval. Currently, three tests are used in the VDT:

- Sensor Range Test (SRT)
- Neighboring Vehicle Test (NVT)
- Neighboring Surface Station Test (NST)

The SRT looks for observations that fall outside of the known sensor range according to hardware specifications. This is the only test available for vehicle observations, excluding air temperature and barometric pressure, because the other vehicle observations (e.g. wiper status) have no ground truth to base QCh tests on. The bounds for the SRT are given in Table 2.

Table 1: List of vehicle observations and associated SRT bounds

Observation	Bounds
Air Temperature	[-40,151°C]
Barometric Pressure	[580,1090 mb]
Vehicle Speed	[-327.65, 327.65 m/s]
Brake Status	[0000, 1111] bits: all off, right rear active, right front active, left rear active, left front active, all on
Brake Boost	[0,2]: not equipped, off, on
Wiper Status	[0,5] and 255: not equipped, off, intermittent, low, high, washer, automatic present
Traction Control	[00, 11] bits: not equipped, off, on, engaged
Stability Control	[00, 11] bits: not equipped, off, on, engaged
ABS	[00, 11] bits: not equipped, off, on, engaged
Headlights	[0000-0000, 1111-1111] bits: 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
Yaw Rate	[0,655.35°/s]
Latitudinal Horizontal Accel	[-20,20 m/s^2]
Longitudinal Horizontal Accel	[-20,20 m/s^2]
Steering Angle	[-655.36, 655.36°]
Steering Rate	[-381,381°/s]

The NVT compares the given vehicle observation to neighboring vehicles on the same road segment. Specifically, the standard deviation and mean of the observations along a 1-mile road segment during the 5-minute VDT snapshot are taken, and then each observation is checked to assure it falls within a certain number of standard deviations of the mean of the road segment. Currently, the VDT uses a threshold of 2.5 standard deviations. This value was chosen based on previous tests of the first set of probe data tested with the VDT, which determined 2.0 was too strict. Sensitivity tests on this threshold were performed and their results given in Section 3.2. With the DTE10 data, the NVT is less discriminating than it would be in larger datasets, because only 9 vehicles, or sometimes less, were present in the entire testing area. Currently, there is no minimum number of observations per road segment required to determine if the standard deviation is meaningful enough for a QCh test.

The NST compares data with the closest surface ASOS station in space and time. The nearest stations are defined as being within a 50-km radius and within 5 minutes of the vehicle observation. If more than one ASOS station meets these criteria, a mean of those observations is taken. A temperature observation passes if it is within 2°C of the ASOS station observation (see Section 3.2 for sensitivity tests). This test is not currently performed on pressure data in the VDT. Vehicle-collected pressure is a station pressure, whereas the ASOS observations being received by the VDT report pressure as reduced to mean sea level (in the Detroit area, this tends to be about a 30-mb difference in average conditions). It is planned to address this issue in the next version of the VDT. For this study, the NST was performed manually using a 10-mb threshold and the WXT520 sensor set up at the test site, which recorded station pressure. For the DTE10 temperature data, the nearest ASOS station was most often Detroit Metro (KDTW). Although the WXT520 was available for a closer comparison during the DTE10 testing, it was decided to retain the original ASOS-based QCh for this test for temperature, because future implementation of the VDT in real time on a nation-wide basis would not have the benefit of such a close sensor. Using the ASOS better allows the QCh analysis in this report to represent future implementation of the VDT. The WXT520 was used for pressure because its observations were reported as station pressure and there were concerns on the accuracy of reducing the vehicle-measured pressures to mean sea level with incomplete elevation and humidity knowledge. Additionally, the VDT assigned WXT520 pressures to the vehicle observations automatically, and ASOS station pressure would have to be obtained and assigned to those observations manually.

After each observation was run through the three tests, a final QCh flag was assigned, which is termed the Combined Algorithm Test (CAT). Development of the confidence levels for this test is ongoing. For this study, three confidence levels were assigned:

- High – observation passes all three QCh tests
- Low – observation passes the SRT but fails either the NVT, NST, or both
- No – observation fails all three tests

Note that the CAT does not dispose of any observations, but merely flags them. This allows the quality of all observations run through the VDT to be examined, keeping in mind their assigned confidence.

3. RESULTS

Once each vehicle observation was run through the VDT with QCh tests applied, three analyses were performed. First, QCh pass rates were examined to see if any meteorological (temperature, wind direction and speed, precipitation) or non-meteorological (day, time of day, vehicle speed) factors affected the pass rates. The DTE10 dataset also retains the vehicle identification, so stratification by vehicle was possible. Second, sensitivity tests were performed to examine how altering the QCh bounds affects pass rates and statistics. Third, the QCh-ed observations, both high and low confidence, were examined in the same manner as QCh pass rates to determine the effects of different factors on the statistical comparisons between the vehicle observations and the WXT520.

3.1 QCh Analysis

For the QCh currently implemented in the VDT, only temperature and pressure are evaluated beyond their sensor range and are presented in this section. SRT pass rates for other vehicle observations are found in Table 2. All observations that did not pass the SRT were those reported as missing values by the vehicles. Brake boost and wiper status were not reported by the Ford Edge vehicles, which made up about one third of the observations. The overwhelming majority of traction control, ABS, headlights, steering angle, and steering rate observations passed the SRT. Failures were reported as missing values by the vehicles, often the first observation in the OBE log.

Table 2: SRT pass rates for vehicle observations excluding temperature and pressure.

Observation	Percent Passed
Vehicle Speed	100.00
Brake Status	100.00
Brake Boost	66.29
Wiper Status	66.29
Traction Control	99.93
Stability Control	99.93
ABS	99.93
Headlights	99.99
Yaw Rate	100.00
Latitudinal Horizontal Accel	100.00
Longitudinal Horizontal Accel	100.00
Steering Angle	99.99
Steering Rate	99.97

Figure 1 gives the percentage of observations passing each QCh test for air temperature and pressure. All observations passed the SRT. With the CAT's current setup, this results in 100% of observations having at least low confidence. Nearly all passed the NVT (99.84% for temperature, 99.82% for pressure). The NST was more discriminating and had the largest effect on which observations were assigned high confidence. Most temperature observations passed (91.66%) while only about one third of pressure observations passed (31.50%). Pressure was reported in a coarse 10-mb resolution, which may have contributed to low pass rates. However, many pressure observations were further off the WXT520 observations than a 10-mb resolution accounts for (see Section 3.3). There is not enough controlled information to separate out whether the poor pass rates are related to collecting pressure on a mobile platform, poor sensor quality/procedure in deriving the barometric pressure from the vehicles' Manifold Absolute Pressure (MAP) and Mass Air Flow (MAF) systems, or both. The percentage of temperature and pressure observations given a high confidence was 91.54% and 31.49% respectively. Because 100% of observations passed the SRT, further stratifications were performed with only the NVT, NST, and CAT high confidence. For the NVT, all stratifications showed no trend (with all pass rates >99%). This caused the NST and CAT high confidence stratifications to show the same trends, so only the results of the NST are shown here.

To determine if pass rates were affected by meteorological conditions, the results of the QCh tests were stratified by temperature, wind direction and speed, and precipitation condition. Temperature and wind observations were supplied by the WXT520 sensor. Precipitation condition was inferred from vehicle wiper status, because the WXT520 does not measure frozen precipitation, and the ASOS station was deemed too far away to be representative of the DTE precipitation condition.

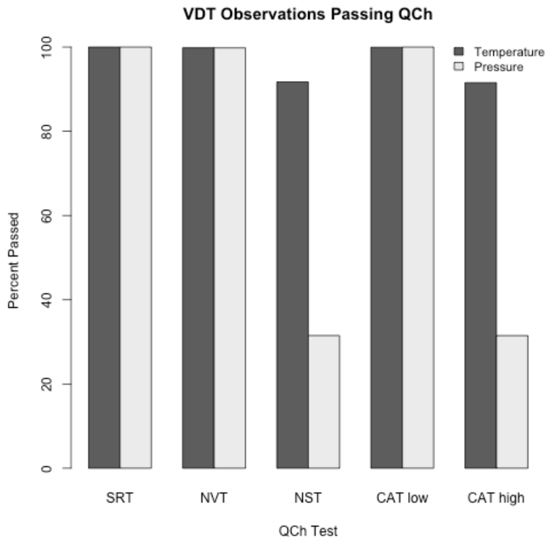

Figure 1: Overall QCh pass rates for temperature and pressure.

Pass rates for both temperature and pressure did not seem to be affected by air temperature or wind direction for any tests (not shown). Pass rates for both temperature and pressure increased slightly with higher wind speeds (Fig. 2). The increase is modest for pressure (~5%) and slightly larger for air temperature (~10%). For precipitation, there did not appear to be much of a trend in pass rates except for pressure, which dropped dramatically (from 43% to 6%) for the “steady” wiper category (Fig. 3). This could be due to sample size issues – there were only 1,326 observations for the “steady” category compared to 118,067 and 39,618 for the “off” and “intermittent” categories respectively. It could also be due to an effect of precipitation on derivations of barometric pressure from the vehicle sensors. Overall though, it appears that meteorological conditions do not have much effect on pass rates. However, the pressure pass rates in heavier precipitation should be kept in mind, as this could impact the usefulness of vehicle pressure observations in precipitating environments if found to be consistent between different datasets.

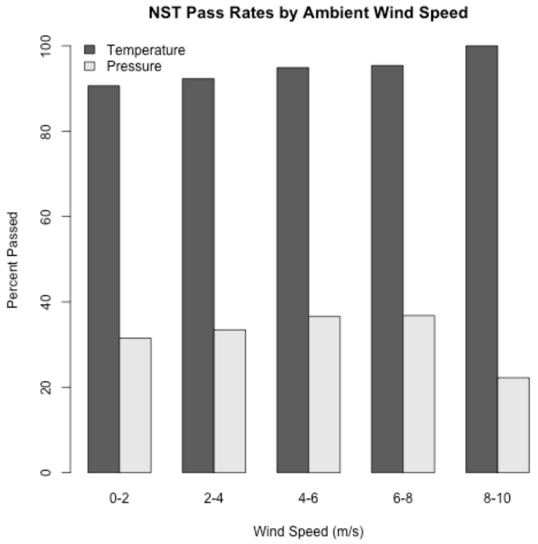

Figure 2: NST pass rates broken down by wind speed.

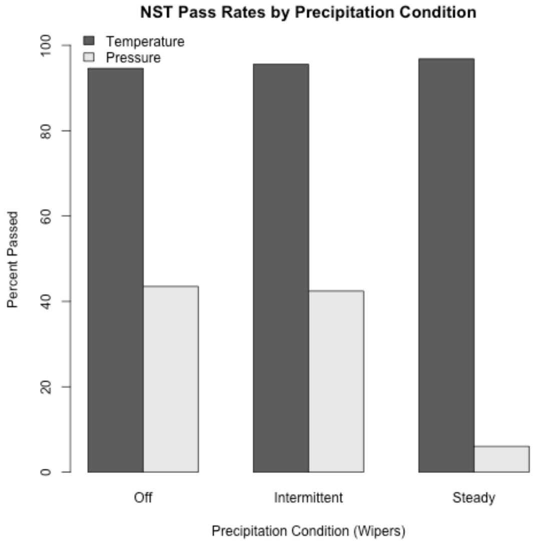

Figure 3: NST pass rates broken down by wiper status.

Non-meteorological factors were also considered and included the following: day, time of day, vehicle speed, and vehicle. Pass rates differed by day (not shown), but there was no exact pattern. For time of day, there was a slight downward trend in both temperature and pressure pass rates from morning to evening, but this trend was fairly weak and inconsistent (not shown). There was also very little difference in pass rates for differing vehicle speeds (not shown). Overall, the additional non-meteorological factors appeared to have little impact on QCh pass rates.

The largest differences in pass rates occurred when the data was stratified by vehicle (Fig. 4). Temperature pass rates were mostly comparable between the vehicles, although the Fords had slightly lower mean NST pass rates than the Jeeps (85.96% compared to 94.77%). For pressure, both e2 and e3 had 0% pass rates for the NST while e4 had only a 23.78% rate. The Jeeps averaged 41.56% of pressure observations passing the NST, although this varied between vehicles from 25.58% for p10 to 55.30% for p8. These results point out a clear need to acquire data from a larger, more variable sample of vehicle types to assess how different makes/models influence pass rates and overall vehicle data quality.

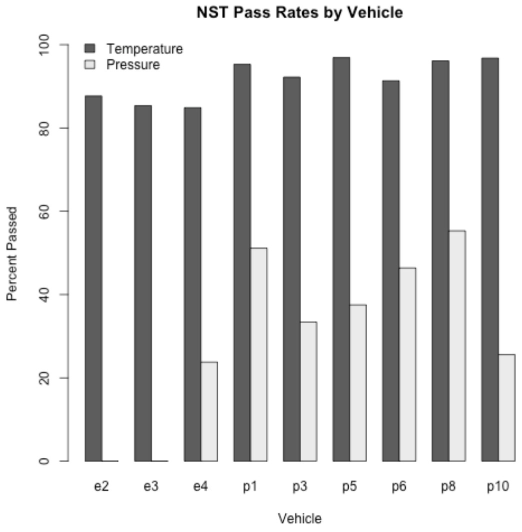

Figure 4: NST pass rates broken down by vehicle. Ford Edges are “e” vehicles and Jeep Grand Cherokees are “p” vehicles.

3.2 QCh sensitivity tests

To assure that the method of QCh selected did not affect the resulting statistics significantly compared to another QCh option, the DTE10 data was run through the VDT four times, each time altering the QCh bounds slightly. For the NST, these tests were run manually for pressure using the WXT520 as described in Section 2.

Pass rates for the different thresholds are presented in Fig. 5. Except for the strictest threshold of 1 standard deviation, there is a nearly 100% pass rate for both temperature and pressure when considering the NVT. For the NST, pass rates increase with a laxer threshold as would be expected. For temperature, except for the strictest threshold, this increase is modest. For pressure, the increase in pass rates is much more pronounced, with only the largest threshold of 20 mb producing a pass rate of higher than 60%. This rate is still less than that for temperature at the strictest threshold, demonstrating the superiority of the temperature measurements to the pressure when QCh-ing against a neighboring surface station.

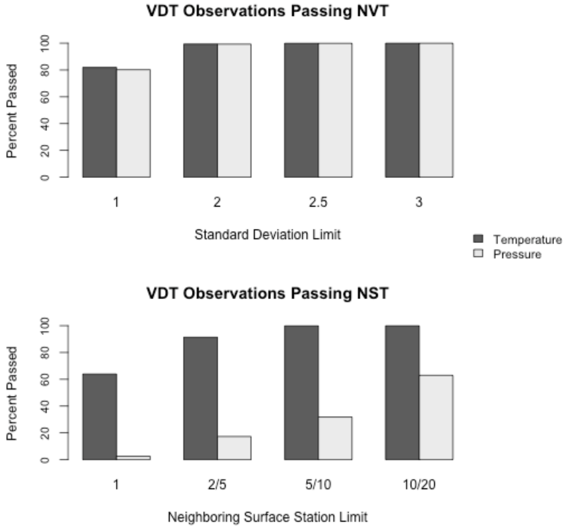

Figure 5: QCh pass rates for temperature and pressure with various thresholds for the NVT and NST.

The NVT was run on both temperature and pressure for standard deviations of 1, 2, 2.5, and 3. In addition to these categories, statistics were run on the data when the test was not applied (e.g. raw data coming out of the VDT). The statistics used were bias, mean absolute error (MAE), and correlation. The bias indicates how far over or under an observation is in relation to another like observation, MAE is used to show how close the measurement of a variable is to its comparison observation, and correlation quantifies the linear relationship between the two variables. For both temperature and pressure, varying the bounds of the NVT did not significantly impact the statistics (not shown).

The NST was run for bounds of 1, 2, 5, and 10°C for temperature and 1, 5, 10, and 20 mb for pressure. With temperature, there was an obvious increase in MAE with laxer thresholds compared to the NVT (Fig. 6). However, these values are all below 1°C. Correlation and bias change little. There is a much larger impact on the pressure statistics (Fig. 6). The bias and MAE degrade to over 20 mb with no QCh, and correlation decreases from 1 to 0.32.

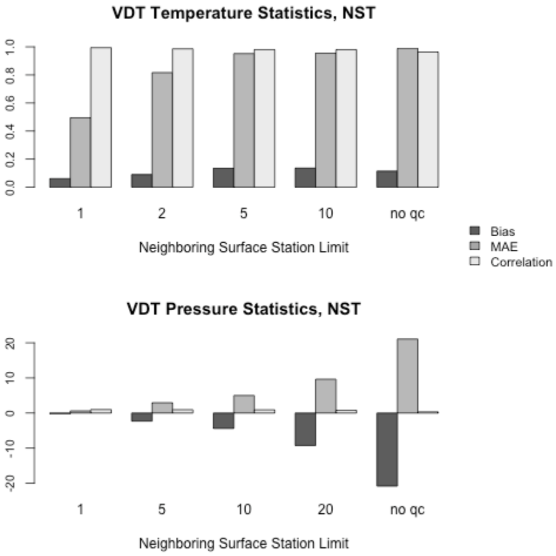

Figure 6: Statistics for temperature and pressure for various thresholds of the NST.

Overall, varying the bounds on the QCh tests for temperature did not significantly affect the statistics. For pressure, varying the NVT had little effect, but varied NST bounds produced large differences in both pass rates and statistics. Part of this may be a result of the 10-mb reporting resolution. Given a future finer reporting resolution, these tests should be rerun to gain a better understanding of the sensitivity of these pressure data to varied QCh thresholds. For the post-QCh analysis, low confidence pressure values are included so that such QCh effects do not produce misleading results.

3.3 Post-QCh analysis

Both low and high confidence data were analyzed to examine the accuracy and bias of the QCh-ed vehicle observations from DTE10, using the WXT520 as truth. Statistics used are bias, mean absolute error (MAE), and correlation (described in Section 3.2).

For a first step, the Student's T-Test for paired observations was performed. The p-value was less than 0.01 for both high confidence temperature and pressure, meaning the vehicle and WXT520 datasets are statistically significantly different. However, the actual difference in the means was small (-0.21°C and -4.33 mb) and the DTE10 variance was within 2 of the WXT520 variance for both observations. Additionally, the difference in medians was relatively small (1.2°C and 7.5 mb), particularly given the vehicles' reporting resolutions of 1°C and 10 mb. Given these results, although there is a statistically significant difference between the vehicle observations and the WXT520, this difference is not physically or practically meaningful. The statistical difference is most likely due to the large sample size.

Overall, the high confidence QCh-ed vehicle observations show favorable comparison to the WXT520 (Fig. 7). This should be expected for pressure, as the WXT520 was used for the NST, but for temperature the NST was run using the KDTW ASOS, and the overwhelming majority of observations were given high confidence. Additionally, low confidence (for DTE10, all observations) temperature statistics also show a relatively close relationship between vehicle and ground sensor observations. Low confidence pressure statistics, like the pass rates presented in the previous subsection, are poor.

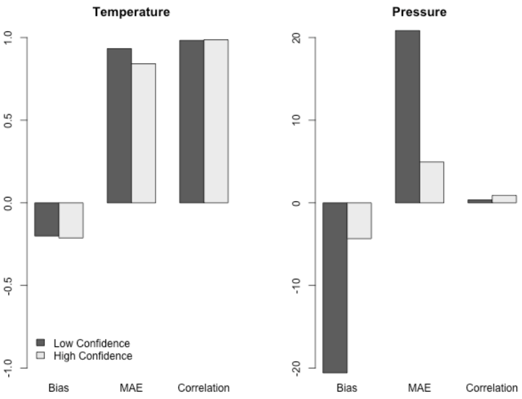

Figure 7: Overall statistics for low and high confidence temperature and pressure observations.

As with the QCh analysis, post-QCh observations of temperature and pressure were stratified to determine any possible effects of meteorological and non-meteorological factors on the results. These factors are the same as used for the QCh results in Section 3.1. Beginning with the ambient temperature, air temperature observations show a bias that tends to be slightly positive below 0°C and slightly negative above 0°C (not shown). MAE decreases slightly with increasing temperature, and correlation increases with temperature. For pressure, the magnitude of bias decreases slightly with increasing temperature and MAE decreases correspondingly (not shown). However, this trend is only evident with the low confidence observations; the high confidence observations change little. For both confidence levels, correlation increases with increasing temperature. For wind direction, temperature bias tends to be more positive with a southerly wind and more negative with a northerly wind (Fig. 8). MAE and correlation vary little with wind direction. Regarding pressure (not shown), the statistics vary between wind directions, but the differences are not large and there is no clear pattern. The bias pattern for temperature could possibly be due to the less than ideal location of the WXT520 in an urban environment, with surrounding trees and buildings affecting temperature readings depending on the wind direction. By wind speed, temperature bias (not shown) is slightly negative for lower speeds and slightly positive for higher speeds, but these differences are less than 0.5°C. There is very little difference in MAE and correlation. For pressure (not shown), the high confidence bias becomes slightly more negative and the MAE increases slightly with faster wind speeds. Correlation varies as well, but not with a consistent trend. Using wiper status to infer precipitation condition, temperature bias varies between conditions but without a clear trend (not shown). MAE improves with increasing wiper rate, and correlation varies little. For pressure, there are no major differences in the bias or MAE between wiper rate categories. In correlation, a slight upward trend is seen with increasing precipitation for low confidence, whereas a slight downward trend is seen with high confidence.

Non-meteorological factors were also examined. By date, there is great variability in the statistics. For temperature, the bias moves from positive values to negative values as the days progress (Fig. 9). A similar trend was seen when moving from cooler to warmer temperatures. This bias pattern may be due to the progression from winter to spring seasons. MAE and correlation vary between days, but without consistent trend (not shown). Pressure statistics vary between days, but there are no obvious trends, and in particular bias and MAE vary little (not shown). By time of day, temperature MAE and correlation varied little apart from the earliest and latest hours, the poorer statistics at these times likely due to fewer observations being available. Temperature bias starts much more negative and moves towards positive values through the day (not shown). However, the actual difference between the biases at various hours is less than 1°C. Pressure statistics exhibit no obvious trend. For vehicle speed (not shown), both temperature and pressure vary little.

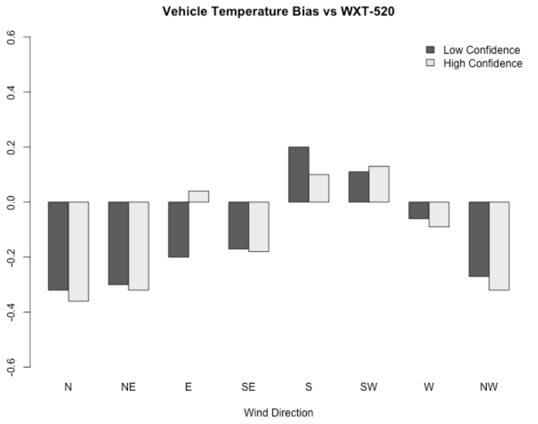

Figure 8: Bias for low and high confidence temperature broken down by wind direction.

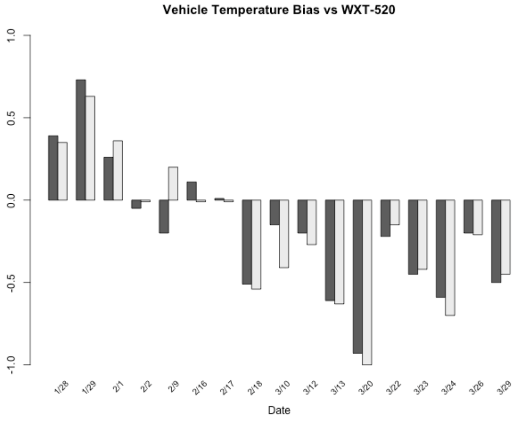

Figure 9: Bias for low and high confidence temperature broken down by date.

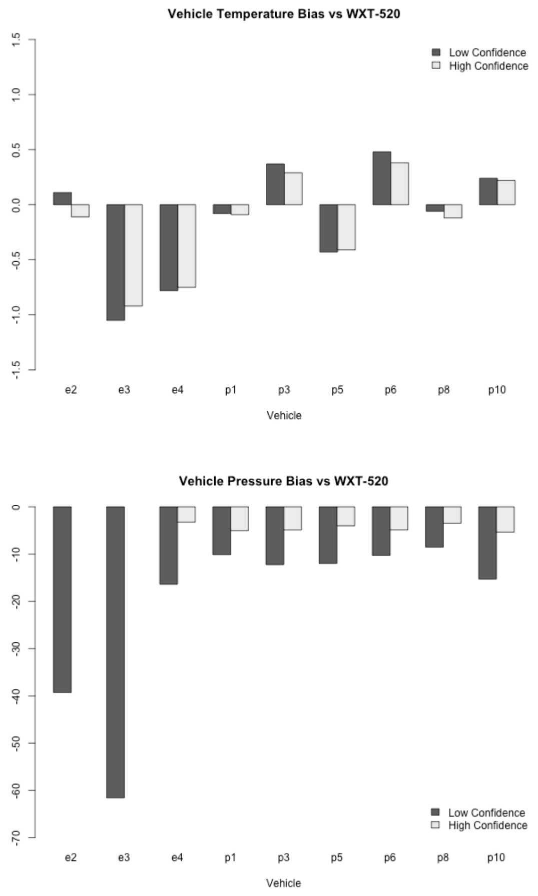

Figure 10: Bias for low and high confidence temperature (top) and pressure (bottom) broken down by vehicle. Ford Edges are “e” vehicles and Jeep Grand Cherokees are “p” vehicles.

When stratifying the statistics by vehicle (Fig. 10 for bias), large differences between each are seen, particularly with bias. The temperature bias shows variation not only between make of vehicle (Ford Edge vs Jeep Grand Cherokee) but also between vehicles of the same make/model. The MAE (not shown) shows less variation, with the Fords having a slightly higher MAE than the Jeeps. Correlation is high among all vehicles. For pressure, the Fords have very high negative biases (Fig. 10) and MAE values (not shown) compared to the Jeeps, particularly e2 and e3, which have biases of -39.29 mb and -61.58 mb respectively for low confidence. These numbers clearly show why no pressure observations from these vehicles passed with a high confidence. Statistics for e4 are more in line with the Jeeps, although it has a more negative bias, higher MAE, and lower correlation than the Jeeps. The statistics vary between Jeeps as well, but not significantly. The differences in pressure statistics are mostly seen with low confidence observations, while high confidence pressure observation statistics vary little.

4. SUMMARY AND CONCLUSIONS

Overall, the vehicle temperature measurements show reasonable agreement with the WXT520. The low and high confidence results for temperature did not differ significantly from each other, implying that the QCh process did not have a large effect on this conclusion. Additionally, varying the QCh bounds for temperature did not significantly affect the statistics. The temperature statistics were also not significantly impacted by different categories of meteorological and non-meteorological factors. There were some differences and trends, but the magnitudes of these were small (within about 1°C) and would likely have little impact in an operational environment.

The pressure comparisons were not as favorable. Low and high confidence results differed greatly, demonstrating that the QCh process would have a large impact on this dataset. Varying the QCh bounds also had an impact on the pass rates for pressure. As with temperature, there were some impacts of meteorological and non-meteorological factors on the statistics, but none had a particularly obvious trend. The largest differences lied between make and model of vehicle, emphasizing the need to test with as many makes and models as possible. Part of the issue with pressure measurements could be due to the coarse 10-mb resolution. Until a more practical resolution is achieved, the extent to which it affects the statistics remains unclear.

These results support the feasibility of collecting air temperature observations from a mobile platform, in particular from ordinary passenger vehicles. These data could prove useful for a variety of applications, both related specifically to road weather and outside this scope, such as model data assimilation. Barometric pressure measurements from these DTE10 vehicles are not useful in their current form. More work must be done to improve their quality. A first step in this work should be a finer reporting resolution. Once this is achieved, a similar analysis to the one presented here can determine whether mobile pressure measurements are feasible with current collection methods.

5. ACKNOWLEDGEMENTS

This research was funded by the Federal Highway Administration. Development/support of the VDT included contributions from Elena Schuler, Gerry Wiener, and Brice Lambi of NCAR.

6. REFERENCES

Drobot, S.D., and Coauthors, 2009: IntelliDrive(SM) road weather research & development – the Weather Data Translator. *Proc. Intelligent Transportation Society of America Annual Conf.,* National Harbor, MD. Intelligent Transportation Society of America, 13 pp.

National Weather Service, cited 2010: Weather Fatalities. [Available online at http://www.weather.gov/os/hazstats.shtml]

Noblis: Eleven-year averages from 1995 to 2005 analyzed by Noblis, based on NHTSA data. [Available online at http://ops.fhwa.dot.gov/weather/q1_roadimpact.htm]

Vaisala, cited 2010a: Vaisala Surface Patrol HD Pavement Temperature and Humidity Sensor Series DSP200. [Available online at http://www.vaisala.com/en/products/surfacesensors/Pages/DSP211.aspx]

Vaisala, cited 2010b: Vaisala Weather Transmitter WXT520. [Available online at http://www.vaisala.com/en/products/multiweathersensors/Pages/WXT520.aspx]