Variability of the Measurement of Temperature in the Canadian Surface Weather and Reference Climate Networks

Gary Beaney, Tomasz Stapf, Brian Sheppard

Meteorological Service of Canada
4905 Dufferin Street, Downsview, Ontario, Canada
Phone: (416) 739-4111, Fax: (416) 739-5721
Gary.Beaney@ec.gc.ca

ABSTRACT

To assess the uncertainties of temperature and humidity measurement, the Test and Evaluation Division of the Meteorological Service of Canada organized a field evaluation of the main temperature and humidity sensor models currently used in Canadian surface weather and climate networks. The field study began in the fall of 2002 at the Centre for Atmospheric Research Experiments (CARE) in Egbert, Ontario. Twelve pairs of temperature and humidity sensors were set up in various configurations with respect to aspiration and shield type. Data was stratified by temperature. Operational comparability and functional precision values were derived to determine the differences between configurations and the variability within configurations.

INTRODUCTION

Over the past few decades, the progressive automation of the Canadian surface weather and climate networks has been undertaken by each of five national regions (Pacific and Yukon, Prairie and Northern, Ontario, Quebec, Atlantic). As a consequence, these networks are populated with a variety of different sensors installed in various configurations. Although individual manufacturers provide performance specifications for these sensors, the Meteorological Service of Canada had not undertaken nation-wide field evaluations.

To assess measurement uncertainties of temperature and humidity, the Test and Evaluation Division of the Meteorological Service of Canada initiated a field evaluation of the major temperature and humidity sensor models currently used in Canadian surface weather and climate networks. The field study began in December, 2002 and continued for a ten-month period at the Centre for Atmospheric Research Experiments (CARE) in Egbert, Ontario. Due to the large number of sensors involved, each parameter (temperature and humidity) was analysed separately While this paper addresses air temperature sensors only, a similar paper presenting the results of the humidity analysis will follow at a later date.

REFERENCE AND TEST SENSORS

Due to the impracticality of representing all sensors and configurations present in the field, this study includes only those most representative of the Canadian surface weather and climate networks. Seven different sensor models were identified as being most representative of those currently in use in the field. For each model tested, two identical sensors were evaluated to establish differences between each other as well as from a common reference.

In addition to evaluating differences in performance between different sensor models, the study also addressed the influence of different sensor configurations. Four different configurations, using different combinations of sensor screen types and aspiration, were established for the majority of sensors tested. The two types of screens most commonly used in the field are wooden Stevenson screens (Figure 1) and Gill screens (Figure 2). These screens are either aspirated (by providing a constant flow of air over the enclosed sensor) or non-aspirated (left to ventilate naturally).

Figure 1 - Test sensors in non-aspirated Stevenson screen.

Figure 2 - Test sensors in Gill 12-plate screens

All sensors in their various configurations included in this study are listed in Table 1. It should be noted that after preliminary analysis, four sensors exhibited uncharacteristically high variations from all other sensors under test (including the reference sensor). As the cause of these variations could not be easily established, the data from these sensors were not included in this analysis, but will be analysed separately at a later date. A list of the sensors removed from this analysis can be found in Appendix A. From this point forward in the paper, any reference to a particular "sensor" encompasses both the sensor model and the installation configuration.

Table 1. Sensors under test and their configurations.

	Sensor	Screen/Shield	Aspiration
CSI 44002A/WS/NA A	CSI 44002A	(WS) Wooden Screen	(NA) Non-Aspirated
CSI 44002A/WS/NA B	CSI 44002A	(WS) Wooden Screen	(NA) Non-Aspirated
CSI 44212/WS/A A	CSI 44212	(WS) Wooden Screen	(A) Aspirated
CSI 44212/WS/A B	CSI 44212	(WS) Wooden Screen	(A) Aspirated
CSI HMP35C/G12/NA A	CSI HMP35C	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP35C/G12/NA B	CSI HMP35C	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP45C/G12/NA A	CSI HMP45C	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP45C/G12/NA B	CSI HMP45C	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP45C/WS/NA A	CSI HMP45C	(WS) Wooden Screen	(NA) Non-Aspirated
CSI HMP45C/WS/NA B	CSI HMP45C	(WS) Wooden Screen	(NA) Non-Aspirated
CSI HMP45C212/G/A A	CSI HMP45C212	(G) Gill	(A) Aspirated
CSI HMP45C212/G/A B	CSI HMP45C212	(G) Gill	(A) Aspirated
CSI HMP45C212/G12/NA A	CSI HMP45C212	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP45C212/G12/NA B	CSI HMP45C212	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP45C212/WS/A A	CSI HMP45C212	(WS) Wooden Screen	(A) Aspirated
CSI HMP45C212/WS/A B	CSI HMP45C212	(WS) Wooden Screen	(A) Aspirated
CSI HMP45C212/WS/NA A	CSI HMP45C212	(WS) Wooden Screen	(NA) Non-Aspirated
CSI HMP45CF/G12/NA A	CSI HMP45CF	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP45CF/G12/NA B	CSI HMP45CF	(G12) Gill 12-plate	(NA) Non-Aspirated
CSI HMP45CF/WS/NA A	CSI HMP45CF	(WS) Wooden Screen	(NA) Non-Aspirated
CSI HMP45CF/WS/NA B	CSI HMP45CF	(WS) Wooden Screen	(NA) Non-Aspirated
CSI PRT1000/WS/NA A	CSI PRT1000	(WS) Wooden Screen	(NA) Non-Aspirated

Figure 3 - Field reference temperature sensor (YSI20048) in aspirated wooden screen.

The average of three reference sensors of the same model type (YSI SP20048) served as the site reference. This sensor is commonly used operationally by the Meteorological Service of Canada and has a long record of good, reliable performance. All three reference sensors were installed in aspirated wooden screens (Figure 3) in a triangle formation at the test site. The average of these three reference sensors was taken to represent the true temperature at the centre of this triangle. Test sensors were installed in a configuration which assured a distance of no more than 10 m from the centre of the reference triangle in accordance with the standard ASTM 4430. All requirements outlined in ASTM 4430 were followed and met for this experiment. The physical layout of all reference sensors and sensors under test is illustrated in Figure 4. A general view of the setup at the CARE test site is provided in Figure 5.

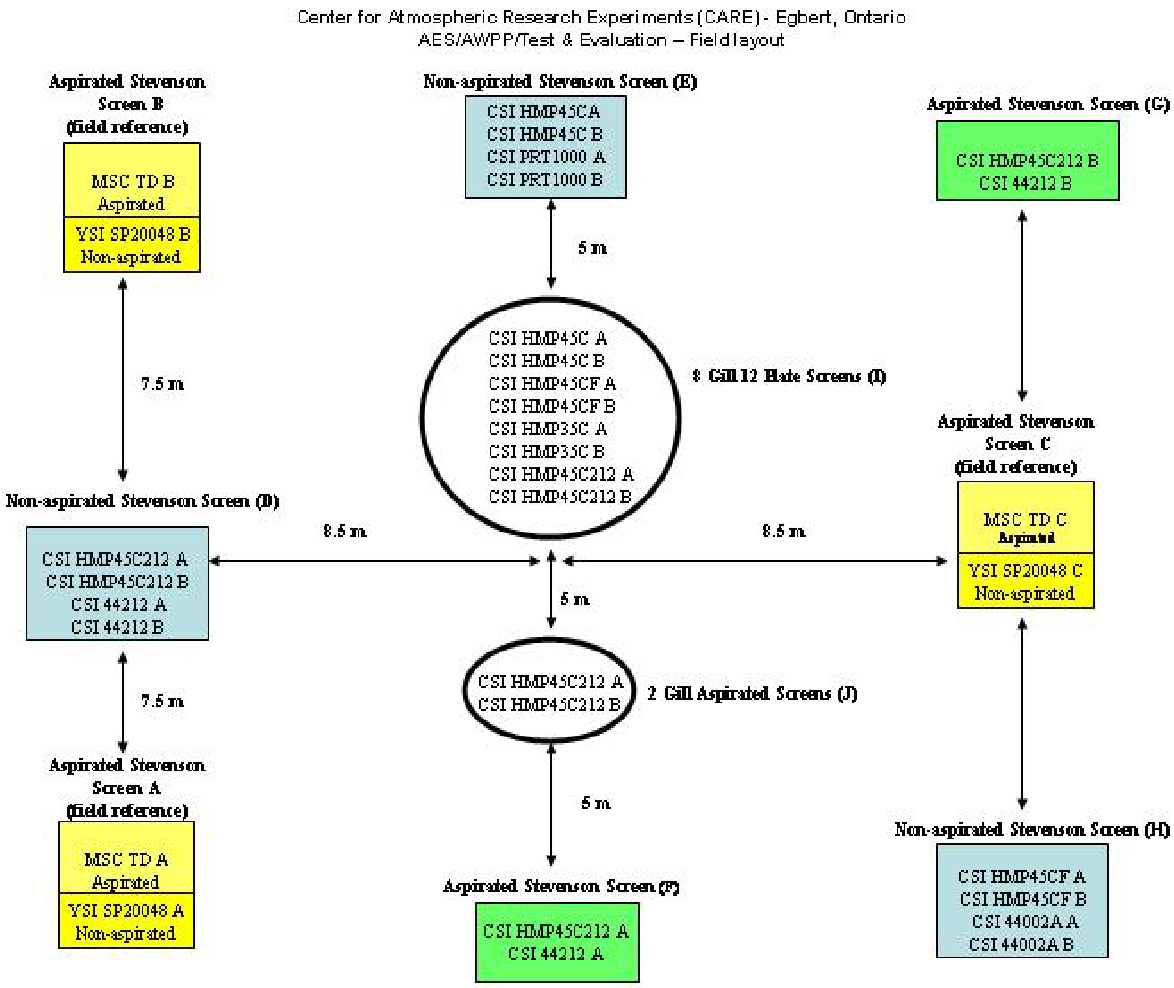

Figure 4 - Layout of all reference and test sensors at CARE test site. Reference sensors are located in the yellow boxes and are labeled (MSC TD A Aspirated).

Figure 5 - General view of temperature/humidity sensors installation in CARE test site.

All data collected for this analysis, from both reference sensors and sensors under test, were of minutely resolution. Minutely reference average values were only included in the analysis when the differences between all reference sensors was less than 0.5^{o}C. To establish a common test period, minutely values were only included in the analysis when data were present for all sensors under test and the reference. If any one sensor was missing a particular minutely value, that minutely value was removed from all other sensors under test.

DATA ANALYSIS

After filtering and the removal of missing data, the resulting dataset was separated into three categories based on reference temperature:

1) <=-5^{o}C
2) >-5^{o}C and <=5^{o}C
3) >5^{o}C

These temperature categories were established to identify temperature-dependent variability in the sensors under test. Each of these three temperature categories was analysed by applying the formulas for operational comparability and functional precision.

Operational comparability was used to quantify the degree of variability of sensors under test from the reference. Operational comparability as defined in ASTM 4430 represents the root mean square of the difference between two simultaneous readings from two systems (reference and sensor under test) measuring the same quantity in the same environment.

Functional Precision was used to quantify the degree of variability of identical sensors under test from each other. Functional precision as defined in ASTM 4430 represents the root mean square of the difference between simultaneous readings from two identical sensors.

RESULTS

Operational comparability scores for all three temperature categories are presented in Figures 6, 7 and 8 (aspirated sensors are highlighted in red). Of all configurations tested, operational comparability scores for aspirated sensors (in both Gill and wooden screens) best agreed with the reference. As the reference average sensors were aspirated, this close agreement was somewhat expected. In addition to aspiration dependence, temperature dependence was also apparent with respect to operational comparability with the lowest scores for most sensors occurring in the middle temperature range (>-5^{o}C and <=5^{o}C).

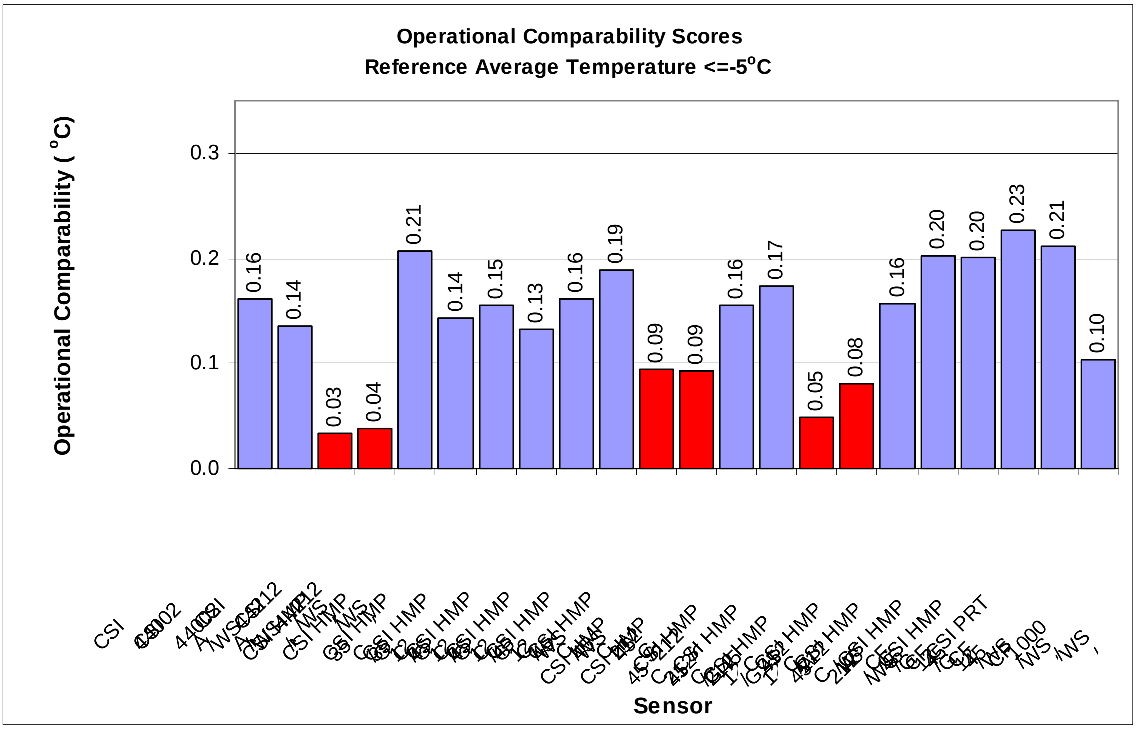

Figure 6 – Operational comparability scores for all sensors under test when reference temperature is <=-5°C. Red bars represent aspirated sensors. Blue bars represent non-aspirated sensors.

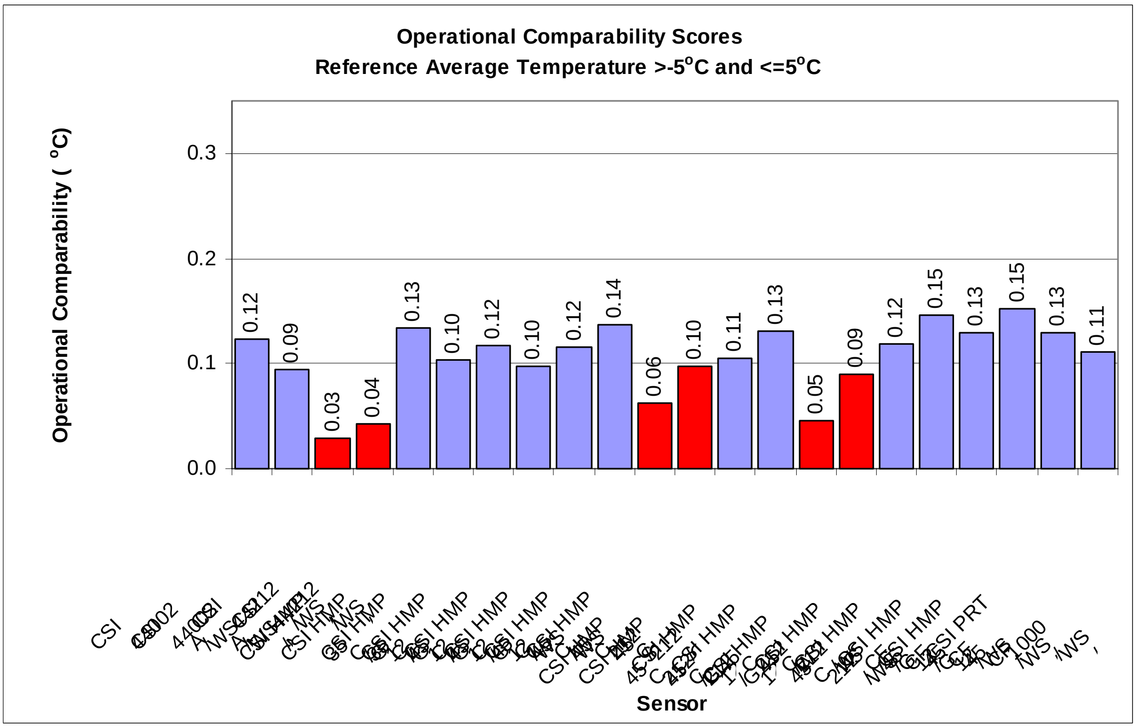

Figure 7 – Operational comparability scores for all sensors under test when reference temperature is >-5°C and <=5°C. Red bars represent aspirated sensors. Blue bars represent non-aspirated sensors.

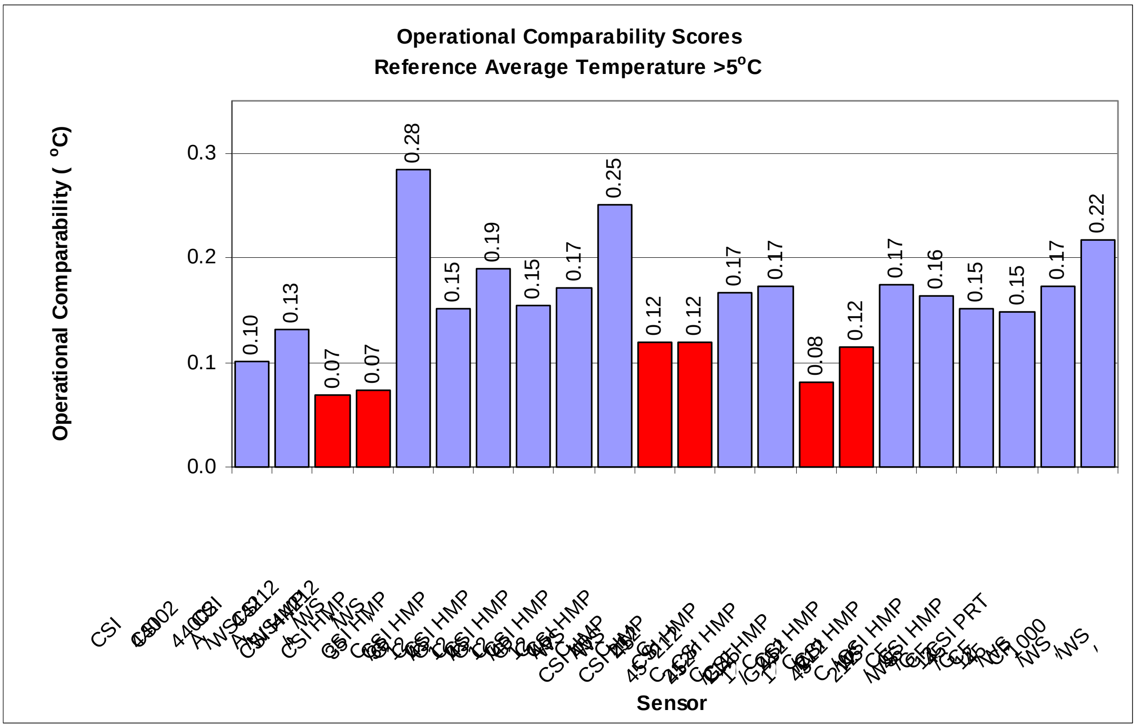

Figure 8 - Operational comparability scores for all sensors under test when reference temperature is >5°C. Red bars represent aspirated sensors. Blue bars represent non-aspirated sensors.

The distributions of the differences (between the reference and each sensor under test) were analysed and student's T-tests were done to determine whether differences between operational comparability scores were a result of higher variability or simply a higher population mean. The results of T-tests comparing the means of all sensors under test and the reference for all three temperature categories are displayed in Table 2. All sensors which have significantly different means from that of the reference are highlighted in red. The majority of sensors in all three temperature categories had significantly different means at the 95% confidence level. Some temperature dependence of T-test results was observed with 6 of 22 sensors not showing significant differences in the <=-5°C temperature category, and only 2 of 22 sensors not showing significant differences from the reference in the >-5°C and <=5°C temperature category. No consistent patterns with respect to sensor model, aspiration, or screen type were observed in the t-test results.

Table 2 – T-test results for comparison of reference and sensor under test means. All sensors highlighted in red were shown to have significantly different means at the 95% confidence leve1.

<=-5°C	>-5°C and <=5°C	>5°C
CSI 44002A/WS/NA A	CSI 44002A/WS/NA A	CSI 44002A/WS/NA A
CSI 44002A/WS/NA B	CSI 44002A/WS/NA B	CSI 44002A/WS/NA B
CSI 44212/WS/A A	CSI 44212/WS/A A	CSI 44212/WS/A A
CSI 44212/WS/A B	CSI 44212/WS/A B	CSI 44212/WS/A B
CSI HMP35C/G12/NA A	CSI HMP35C/G12/NA A	CSI HMP35C/G12/NA A
CSI HMP35C/G12/NA B	CSI HMP35C/G12/NA B	CSI HMP35C/G12/NA B
CSI HMP45C/G12/NA A	CSI HMP45C/G12/NA A	CSI HMP45C/G12/NA A
CSI HMP45C/G12/NA B	CSI HMP45C/G12/NA B	CSI HMP45C/G12/NA B
CSI HMP45C/WS/NA A	CSI HMP45C/WS/NA A	CSI HMP45C/WS/NA A
CSI HMP45C/WS/NA B	CSI HMP45C/WS/NA B	CSI HMP45C/WS/NA B
CSI HMP45C212/G/A A	CSI HMP45C212/G/A A	CSI HMP45C212/G/A A
CSI HMP45C212/G/A B	CSI HMP45C212/G/A B	CSI HMP45C212/G/A B
CSI HMP45C212/G12/NA A	CSI HMP45C212/G12/NA A	CSI HMP45C212/G12/NA A
CSI HMP45C212/G12/NA B	CSI HMP45C212/G12/NA B	CSI HMP45C212/G12/NA B
CSI HMP45C212/WS/A A	CSI HMP45C212/WS/A A	CSI HMP45C212/WS/A A
CSI HMP45C212/WS/A B	CSI HMP45C212/WS/A B	CSI HMP45C212/WS/A B
CSI HMP45C212/WS/NA A	CSI HMP45C212/WS/NA A	CSI HMP45C212/WS/NA A
CSI HMP45CF/G12/NA A	CSI HMP45CF/G12/NA A	CSI HMP45CF/G12/NA A
CSI HMP45CF/G12/NA B	CSI HMP45CF/G12/NA B	CSI HMP45CF/G12/NA B
CSI HMP45CF/WS/NA A	CSI HMP45CF/WS/NA A	CSI HMP45CF/WS/NA A
CSI HMP45CF/WS/NA B	CSI HMP45CF/WS/NA B	CSI HMP45CF/WS/NA B
CSI PRT1000/WS/NA A	CSI PRT1000/WS/NA A	CSI PRT1000/WS/NA A

The most apparent pattern observed, when comparing operational comparability scores and distributions of differences for sensors under test, involved aspirated vs. non-aspirated sensors. The distribution of the differences for the HMP 45C212 WS in aspirated and non-aspirated configurations are displayed in Figures 9 and 10. The aspirated sensor distribution appears more peaked (mesokurtic) while the non-aspirated sensor distribution appears less peaked (platykurtic), and less Gaussian with a skew toward positive differences from the reference average. These differences are common when comparing all aspirated vs. non-aspirated sensors and were exacerbated in the >5°C temperature category with differences in non-aspirated sensors becoming even more pronounced.

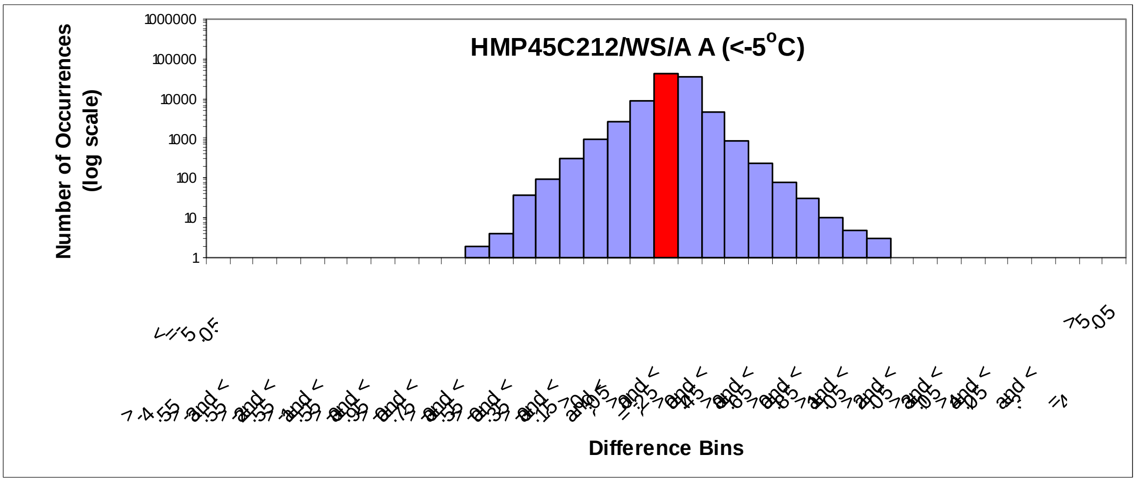

Figure 9 – Example of a frequency distribution of the differences between a sensor under test and the reference in an aspirated screen. The red bar represents the middle (optimum) difference bin.

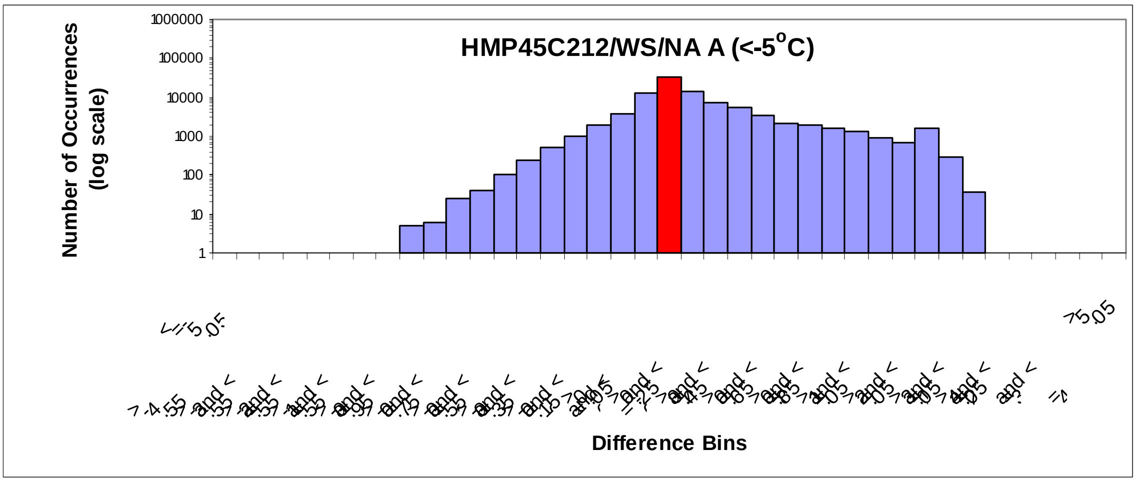

Figure 10 – Example of a frequency distribution of the differences between a sensor under test and the reference in a non-aspirated screen. The red bar represents the middle (optimum) difference bin.

Differences from the reference average for the two sensors with the best and worst operational comparability scores are illustrated in Figure 11 (Note - for graphical purposes, the minutely datasets were reduced in resolution by including only the instantaneous minutely value on the hour. As the summary statistics of mean and coefficient of variation for the minutely and hourly series are virtually identical (Appendix B), the hourly time series provided a rough graphical representation of differences from the reference). As illustrated in Figure 11, both the mean of the differences and degree of variability can differ significantly between sensors under test. Operational comparability scores for all other sensors under test lie between these two extremes.

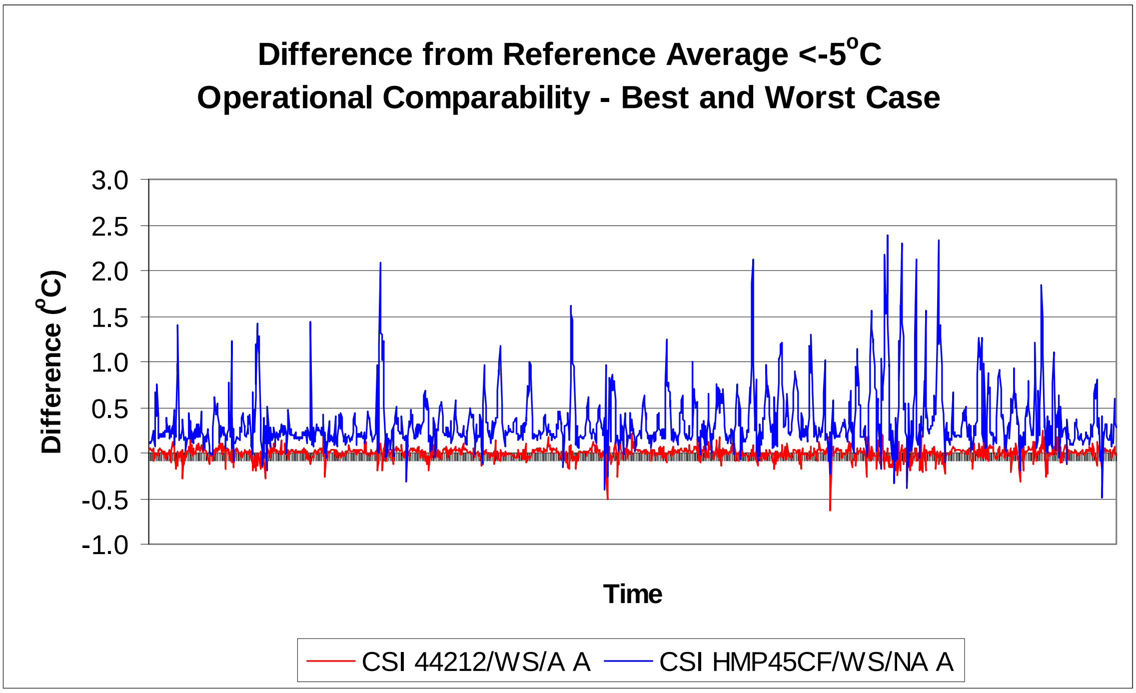

Figure 11 - The two time series represent the differences from the reference average for the sensors with the best and worst operational comparability scores.

While in the <=-5°C temperature category the CSI HMP45CF/WS/NA A sensor had the highest operational comparability score and consequently was shown to compare the worst with the reference, of all the identical pairs of sensors tested, this sensor had the lowest functional precision score. Hourly differences from the reference average for the two identical CSI HMP 45CF/WS/NA sensors (A and B) are illustrated in Figure 12. Hourly differences from the reference average for the pair of identical sensors with the worst functional precision score (CSI HMP35C/G12/NA A) are illustrated in Figure 13.

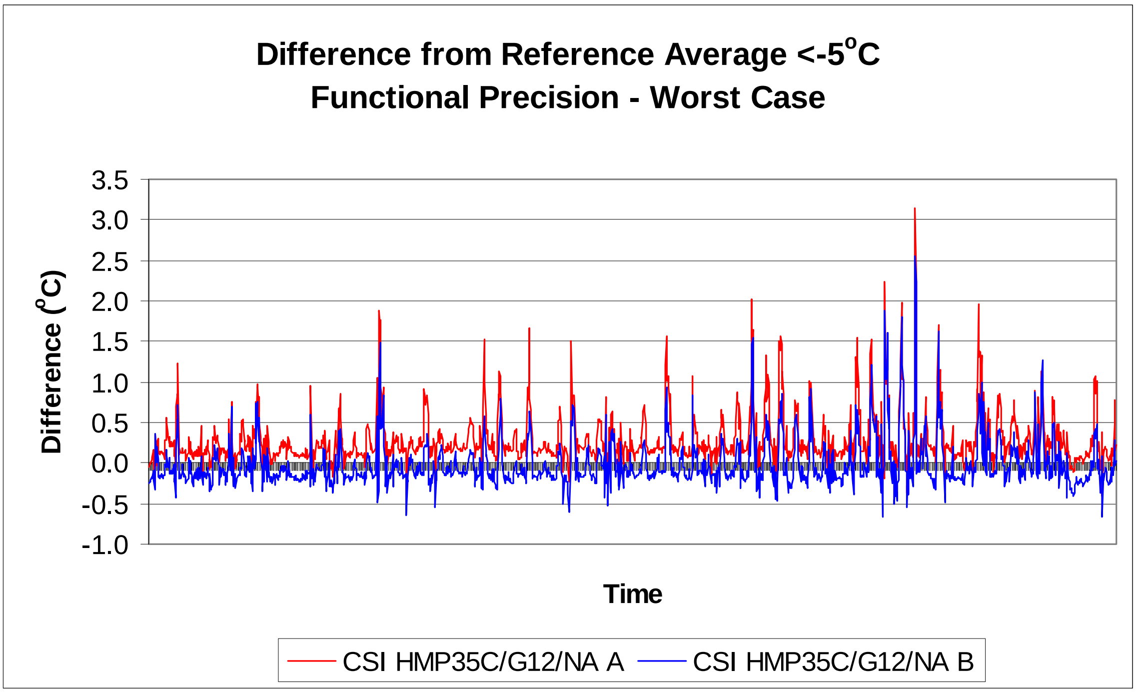

Figure 12 - The two time series represent the differences from the reference average for the pair of sensors with the best functional precision score.

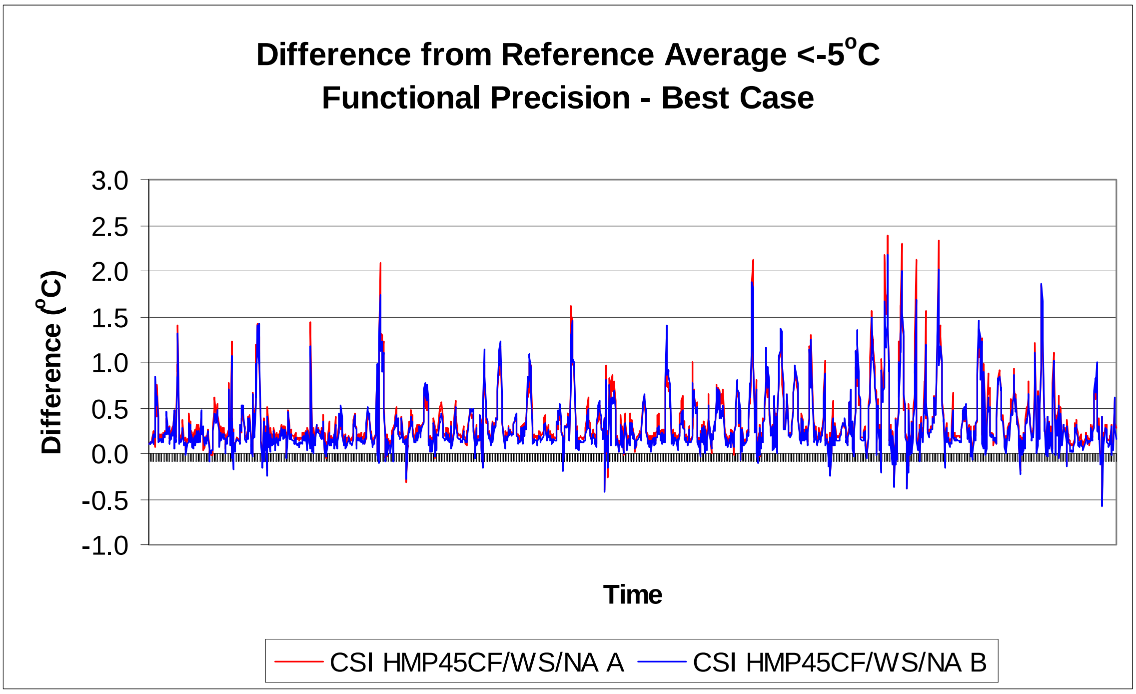

Figure 13 - The two time series represent the differences from the reference average for the pair of sensors with the worst functional precision score.

Functional precision scores for all three temperature categories are displayed in Figures 14, 15 and 16. As was observed with operational comparability, of the three temperature categories, functional precision scores were lowest in the middle category (<-5°C and >=5°C). Unlike operational comparability results, no clear pattern between aspirated and non-aspirated sensors was observed when comparing functional precision scores.

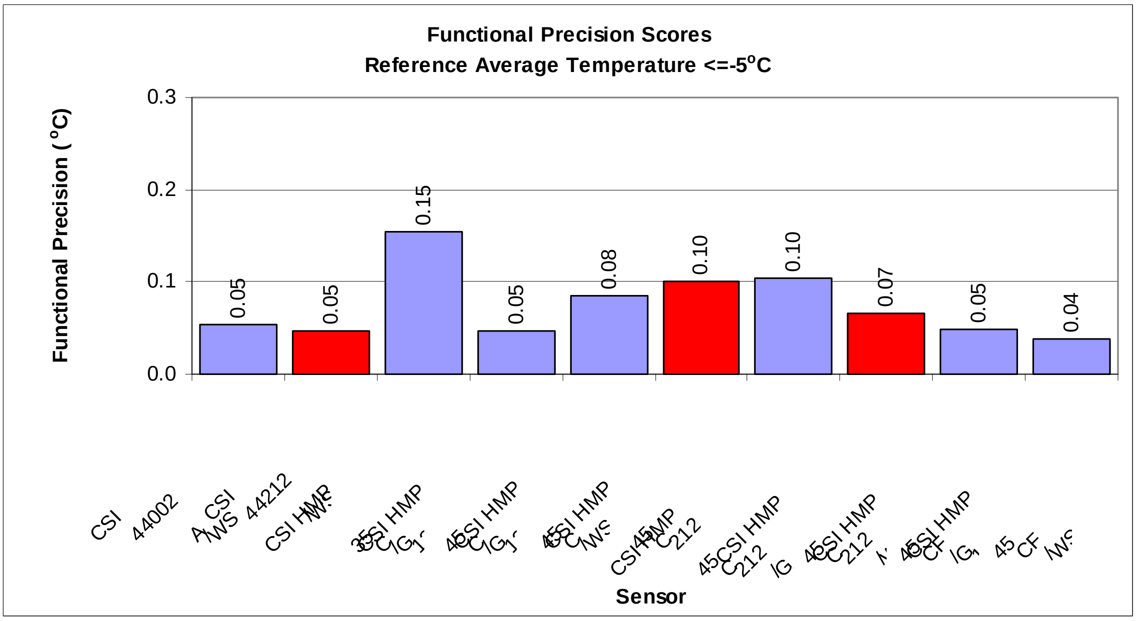

Figure 14 – Functional precision scores for all sensors under test when reference temperature is <=-5°C. Red bars represent aspirated sensors. Blue bars represent non-aspirated sensors.

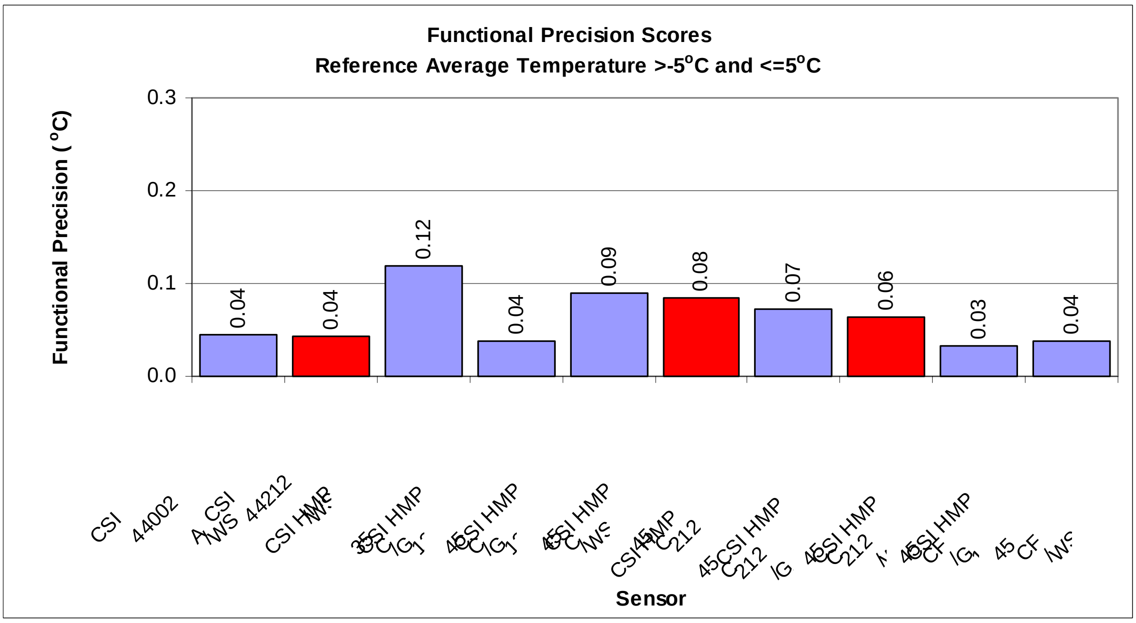

Figure 15 – Functional precision scores for all sensors under test when reference temperature is >-5°C and <=5°C. Red bars represent aspirated sensors. Blue bars represent non-aspirated sensors.

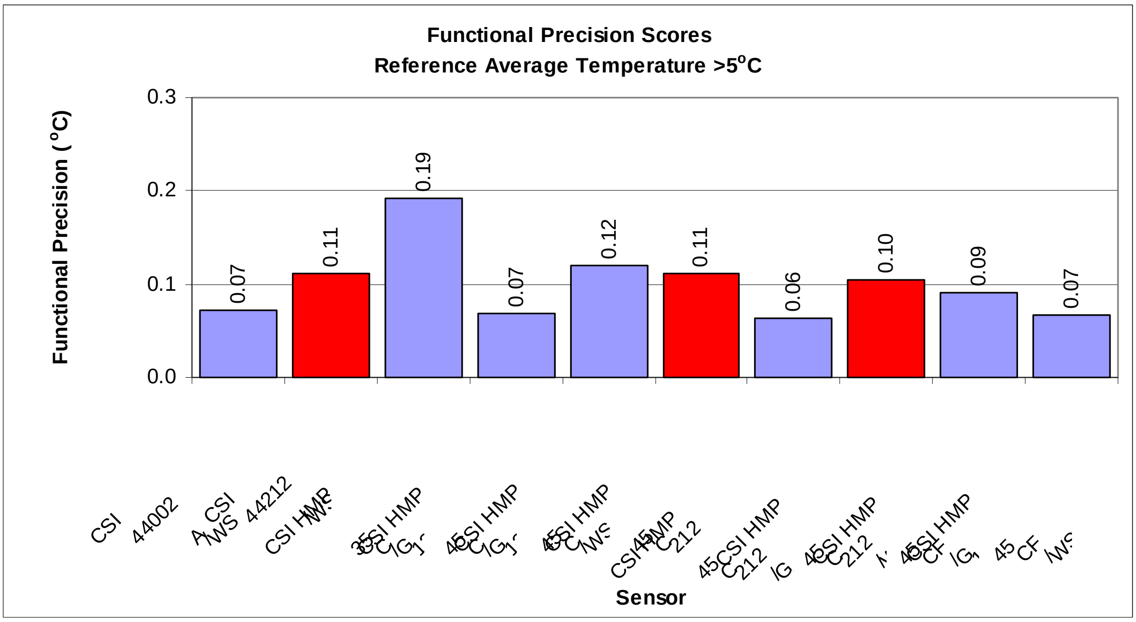

Figure 16 – Functional precision scores for all sensors under test when reference temperature is >5^{o}C. Red bars represent aspirated sensors. Blue bars represent non-aspirated sensors.

DISCUSSION

Both operational comparability and functional precision scores from this analysis illustrate a wide range of variability among operational sensors in the Canadian surface weather and climate networks. Temperature dependence with respect to both the distributions of the differences from the reference and the differences between population means was apparent in both operational comparability and functional precision scores. Further stratification of the data by temperature may better identify the temperature ranges in which particular differences occur. The stratification of datasets by other meteorological parameters such as wind speed may also aid in further identifying the specific differences between operational sensors. The significant differences observed between aspirated and non-aspirated sensors may be less apparent during higher winds due to greater natural ventilation. It should be noted however that the results of this experiment represent a quantification of differences observed in one particular climate regime. The climatology of the test site for this study is not representative of all measurement sites throughout the country. As such, the differences observed between sensors may not be of consistent magnitude or direction in other Canadian climate regimes. To fully illustrate the differences between currently operational sensors, similar experiments would be required in different climate regimes throughout Canada.

The fact that the same sensor (CSI HMP45CF/WS/NA) had the highest operational comparability score and the lowest functional precision score in the <=-5^{o}C temperature category illustrates the need for balance between sensor precision and sensor consistency in a climate network. Ideally, one would prefer a sensor which is shown to be both close to the “truth” and consistent in operation from sensor to sensor. If one is forced to answer the question of whether closeness to the “truth” coupled with unreliability is better or worse than reliability coupled with disparities from the “true” temperature, the answer is not straight forward. However, as systematic differences are traditionally easier to identify and remove than random differences, the latter may be the better choice; as long as the data from such sensors is suitably adjusted prior to use.

CONCLUSION

Due to the vast geographic nature of Canada, the ability to properly identify variations in climate trends depends on instrument continuity from one climate region to another. This experiment identified a wide range of variability among currently operational sensors in the Canadian surface weather and climate networks. The observed variability results from differences in sensor type, sensor configuration and air temperature. Although the Meteorological Service of Canada intends to test possible replacements for operational sensors in different climate regimes throughout Canada, the degree of variability of data residing in the Canadian climate archive from currently operational sensors will remain unknown until similar experiments have been repeated in different climate regimes.

ACKNOWLEDGEMENTS

Our thanks is given to the following colleagues who provided vital contributions to the this project: Peter Bowman, Bob Wilson and George Davis. We are grateful for the fruitful discussions with all of them.

Appendix A

	Sensor	Screen/Shield	Aspiration
CSI 44212/WS/NA A	CSI 44212	(WS) Wooden Screen	(NA) Non-Aspirated
CSI 44212/WS/NA B	CSI 44212	(WS) Wooden Screen	(NA) Non-Aspirated
CSI HMP45C212/WS/NA B	CSI HMP45C212	(WS) Wooden Screen	(NA) Non-Aspirated
CSI PRT1000/WS/NA A	CSI PRT1000	(WS) Wooden Screen	(NA) Non-Aspirated

Sensors removed from experiment prior to analysis due to uncharacteristically high differences from both the reference and all other sensors under test.

Appendix B

	Minutely (N=406299)		Hourly (N=6763)	
Sensor	Mean	Coef.Var.	Mean	Coef.Var.
Reference Average	6.019103	2.15205	6.012748	2.15510
CSI 44002A/WS/NA A	6.124509	2.08976	6.118610	2.09253
CSI 44002A/WS/NA B	6.052633	2.11524	6.046364	2.11813
CSI 44212/WS/A A	5.998492	2.15639	5.992243	2.15943
CSI 44212/WS/A B	6.065454	2.14065	6.058375	2.14394
CSI HMP35C/G12/NA A	6.304771	2.06078	6.297886	2.06402
CSI HMP35C/G12/NA B	6.051136	2.15043	6.043970	2.15381
CSI HMP45C/G12/NA A	6.033197	2.16513	6.026342	2.16854
CSI HMP45C/G12/NA B	6.026166	2.16560	6.018894	2.16920
CSI HMP45C/WS/NA A	6.076401	2.14517	6.068909	2.14861
CSI HMP45C/WS/NA B	6.237260	2.08898	6.229667	2.09226
CSI HMP45C212/G/A A	5.981902	2.16791	5.974659	2.17119
CSI HMP45C212/G/A B	6.132289	2.10982	6.126034	2.11266
CSI HMP45C212/G12/NA A	6.102319	2.13094	6.095518	2.13424
CSI HMP45C212/G12/NA B	6.167084	2.09771	6.160065	2.10098
CSI HMP45C212/WS/A A	6.040655	2.14264	6.034338	2.14561
CSI HMP45C212/WS/A B	6.146586	2.10440	6.139071	2.10767
CSI HMP45C212/WS/NA A	6.151596	2.11030	6.144211	2.11349
CSI HMP45CF/G12/NA A	6.144640	2.09118	6.136420	2.09483
CSI HMP45CF/G12/NA B	6.091604	2.10225	6.084622	2.10542
CSI HMP45CF/WS/NA A	6.118393	2.08970	6.111363	2.09296
CSI HMP45CF/WS/NA B	6.057680	2.10879	6.050991	2.11197
CSI PRT1000/WS/NA A	6.243256	2.08055	6.236481	2.08349