Mark One Lifestyle, Inc.

Golden Setup Collection Tool vs. SpectraWiz

Data Collection

- Date: 2015-04-22
- Ticket: AS-299
- Git Repo: data.mkone.co/var/git/science/vessyl/as-299.git
- Git Branch: master

Analysis

- Ticket: VA-110
- Git Repo: git.mkone.co:vessyl-algorithms/algorithms.git
- Git Branch: va-110

Table of Contents

Revision Control

Revision Number	Revision Date	Notes	Owner
1.0.0	2015-05-12	First draft release	Ehson Ghandehari

Executive Summary

- Qualitatively there is no difference between the data collected by the SW and the GSCT
- A quantitative analysis was not possible since the data collection tools did not recored readings from the exact wavelengths (mismatch in wavelength)
- It was observed that GSCT recorded data in a greater wavelength range than SW, resulting in more data points in extreme sides of the spectrum. However, this issue can be ignored since Science is not interested in data from these regions.

Introduction

Mark One Lifestyle has a StellarNet Silver Nova spectrometers that comes with the SpectraWiz (SW) software. Mark One has developed its own software tool called the Golden Setup Collection Tool (GSCT). SpectraWiz is a mature tool and there was concern that the GSCT may not be collecting and processing the data in the same way. The purpose of this study was to examine these two tools.

Methods

Data Collection

The data was collected under AS-299. The data was collected using a clean, dry cuvette. The cuvette was cleaned three times with distilled water, using a squirt bottle. Then it was rinsed once with isopropyl alcohol (IPA) and blown dried with pressurized air. The clean and empty cuvette was placed in the holder and clamped into position. The mirrored surface was contralateral to the fibre cable. Using first the GSCT, a series of 128 scans, a trial, was collected and stored. After closing the GSCT, without any change to the system, SW tool was launched. A series of 128 scans, a trial, was collected and stored. It must be noted that temperature compensation feature was *activated* in the whole process of AS-299 data collection.

Setup

The setup used in the data collection was the red golden setup. The red golden setup consisted of a light source (Ocean Optics DH-2000; serial number 005400821), an optical fibre (Thorlabs; M200L02S-UV), an optical switch (Ocean Optics; INLINE-TTLS), a custom bifuricated fibre optic reflection probe (Thorlabs; 6 fibres from the light source and 1 fibre to the spectrometer; the length from the light source to the cuvette or the spectrometer to the cuvette is 1 m each; the fibre used in this reflectance probe is a Thorlabs FG550UEC), a custom refurbished cuvette holder (Mark One Lifestyle Inc.), a quartz glass cuvette (Thorlabs; CV10Q3500F) with a custom aluminum mirrored surface, and a spectrometer (StellarNet; Silver Nova model, serial number 15040704). The light source was connected to the optical switch by a M200L02S-UV fibre. The optical switch was connected to the front side of the cuvette holder by one leg (containing 6 fibre) of the reflection probe, and the the other leg (containing 1 fibre) was connected to the spectrometer. The spectrometer was placed in an environmental chamber (Espec BTL-433) at a temperature of 15 C and 70% relative humidity. The spectrometer responded to a wavelength range from 178.28 nm to 1121.7 nm with 2051 individual wavelengths.

The SpectraWiz software was v5.33, and the GSCT software was build number 1.1.5589.42971. The computer running the tool was a Dell laptop running Windows 8.

Data Analysis

In GSCT case, The data was converted from XML format into tidy formatted comma separated values and all trials were concatenated together and compressed. This was done using a custom bash script that calls gsxmltidy.py and the output is concatenated. This output is then compressed using gzip. These compress files were read into R version 3.1.2 (2014-10-31). Within R a four stage cleaning process was performed. All stages are optional and can be performed in any order. They include a conversion phase where the data is converted into the correct data type. This phase is often performed, in part, during the read operation. The transform phase is used to manipulate the values of the data and correct any errors. During the filtering phase, unneeded or erroneous data is removed from the data set. In the final phase the data is transformed into tidy format, if required.

After the cleaning process, the data is run through a series of checks. These checks are independent of the cleaning phase and they test the quality of the data and assumptions that are being made. These would include, but not limited to, checks to see if the data is the correct data type, if there are missing values, the correct number of levels exist in a factor, values are in range and to make sure that all data is present.

The light source produces several very large spikes in the spectrometer data. This causes saturation in some wavelengths and bleed over into the adjacent wavelengths. Therefore, if the reflectance data was clipped, reflectance measures greater than 58890, in any trial it was removed for all trials collected by a given operator. To remove the bleed over, wavelengths that were +/- 1 nm an either side were also removed.

The GSCT AS-299 data were loaded, cleaned and checked.

In SW case, The white, dark, and drink trials were collected as “episodic scans,” feature, creating a file with an EP extension. These files were not readable in TextEditor or MS Office. Therefore, the files were read back in SpectrWiz as a 3D layout of 128 scans. Then each scan was saved as an SSM file (readable by Excel and TextEditor). A function was scripted in R programming language, to extract the data, and concatenat the data into a CSV file. The SW recorded the wavelength range from 183 nm to 1117.5 nm with 1860 individual wavelengths. The SW AS-299 data were loaded, cleaned, and checked.

Results

Figure 1, shows the mean reflectance collected in each software tool from an empty, clean cuvette. The thickness of the line is one standard deviation on either side of the mean. A different colour (red for GSCT, blue for SW, and black for overlap) is used for each tool and a qualitative assessment shows that the signal recorded is in general agreement between the tools. Qualitively, it can be seen that the two lines

overlap in all wavelengths. The discrepancies between GSCT and SW spectra at the extreme sides are due to the larger wavelength recording range of GSCT (GSCT: 178.28 nm to 1121.7 nm vs. SW: 183 nm to 1117.5 nm). There was also a slight difference between two spectra in the Deuterium light saturation region (655-660 nm). It was noticed that GSCT and SW recorded data at different wavelengths. As an example, GSCT collected data at wavelengths 183.3663, 183.7904, 184.2145, 184.6383, but SW collected data at wavelengths 183, 183.5, 184, 184.5. Because of this wavelength mismatch, in the data cleaning process, different wavelengths were filtered out because of signal saturation. This caused the non-overlapping region (655-660 nm range).

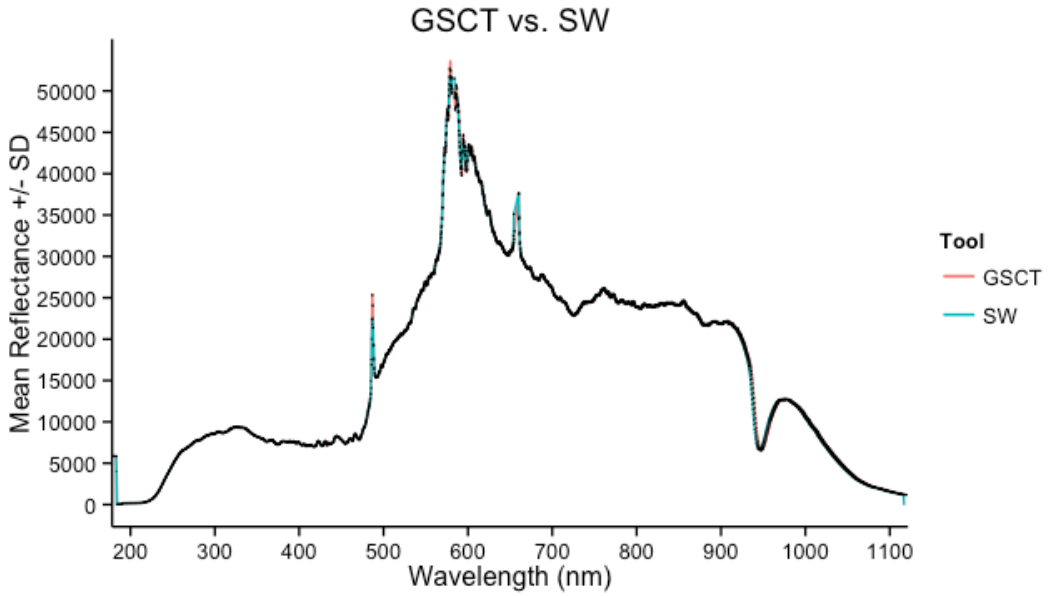

Figure 1: Reflectance for both software tools. The data slightly mismatched in the Deuterium light saturation region (655-660 nm).

As a further study, it would be valuable to compare the difference of the two spectra at each wavelength with a two-sided Student's t-test with equal variances (with hypothesis that there is no difference between the two spectra; this difference is zero). However, because of the reported wavelength mismatch this would not be a reasonable test to run.

Conclusion

This study has demonstrated that qualitatively there is no difference between the data collected by the SW and the GSCT. There were differences between the data in the extreme sides of the spectra (below 183 nm and above 1117.5 nm; the tools

recorded data in different ranges; these regions are not important for Science), and in the 655-660 nm light saturation regions (due to wavelength mismatch). Because of the wavelength mismatch, no quantitative analysis was performed on the data. In case of a request for further quantitative studies, the issue of minor mismatch between data collection tools must be resolved first.