Improving the linearity of infrared diffuse reflection spectroscopy data for quantitative analysis: an application in quantifying organophosphorus contamination in soil

Alan C. Samuels†, Changjiang Zhu§, Barry R. Williams‡, Avishai Ben-David†, Ronald W. Miles, Jr. †, Melissa Hulet‡*

†Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD 21010, §Battelle Eastern Science and Technology Center, 1204 Technology Drive, Aberdeen, MD 21001, ‡ EAI Corporation, 1308 Continental Drive, Abingdon, MD 21009

Theoretical and experimental investigation of the linear relationship between baseline vs. peak intensity in Kubelka-Munk scale and effect of baseline on the reflection scale

Abstract

Both theoretical and experimental investigations were performed to elucidate the linear relationship between baseline vs. peak height in Kubelka-Munk units and baseline offset in reflection log(1/R). The investigations revealed that the baseline offset in log(1/R) space results in the observed linear relationship between baseline and peak height in Kubelka-Munk units.

Introduction

Prof. Griffiths suggested that the linear relationship we observed in Kubelka-Munk (K-M) units between the baseline position vs. peak intensity may actually be an artifact that results when the log(1/R) spectra are converted into K-M units. We undertook the following theoretical and experimental investigations to verify his hypothesis.

Theoretical Study and discussion

A spectrum of three Lorentzian bands was synthesized in log(1/R) scale using Galactic Grams/32 AI to simulate a condensed phase IR absorption feature. As indicated

in Figure S-1a, the spectral region considered was from 0 to 3000 cm^{-1}. The three synthesized bands are located at 1000, 1500 and 2000 cm^{-1} with bandwidth 10, 20 and 6 cm^{-1}, and band intensity 0.5, 0.8 and 0.3 absorption units on the log(1/R) scale. The spectral data interval was 0.2 cm^{-1}/data point. The synthesized bands were created such that there was a data point located at exact peak maximum, and this constraint was preserved through all later data processing.

Six baseline offsets were then added into the synthesized spectrum in order to create six new reflection spectra as shown in Figure S-1b. The offset values were 0.1234487, 0.6834487, 0.8834487, 0.4274487, 0.1974487 and 0.000487 units on the log(1/R) scale. Finally those log(1/R) spectra with baseline were converted into Kubelka-Munk units (see Figure S-1c).

The baseline offset values (B_j) were obtained at 0 cm^{-1} from each of converted Kubelka-Munk spectra. After subtraction of each B_j from its spectrum, $f(R_\infty,\nu)_j$ the value at each peak maximum was extracted for all peaks in the spectra. Figure S-2 shows the plot of $f_0(R_\infty,\nu)_j$ against B_j. The least square linear fit was performed on all three peaks in this plot. Fitting results, R values, indicate that there is indeed a linear relationship between $f_0(R_\infty,\nu)_j$ and the left baseline of peak just as was observed in the experimental data (Figure 1 of the parent article). It is based on this linear relationship that we developed our Kubelka-Munk Correction (KMC) approach, the combination of MSC (multiplicative scatter correction) and BPC (baseline peak correction in the Kubelka-Munk scale) as is given in Eq. 4 of the parent article. Careful observation reveals that it is better to use a nonlinear fit to describe the relationship when B_j approaches 0 value.

We believe that our theoretical investigation demonstrates that Prof. Griffiths' statement is true. That is to say, the linear relationship between baseline offset and peak intensity in Kubelka-Munk units could be the result of the addition of baseline offset in log(1/R) space followed by conversion using the Kubelka-Munk relationship.

Experimental investigation and discussion

A different approach was taken for the experimental study. We corrected the spectra in both log(1/R) space and Kubelka-Munk space, then compared the results in Kubelka-Munk space. MSC was applied in both cases to correct for random ~~scatter~~ spread in the raw data.

We selected seven measurements of caffeine in KBr at a concentration of 10.34 mg/g and six measurements of EMPA in soil/KBr at a concentration of 0.797 mg/g for this work. The instrumental conditions used to acquire these spectra were the same as described in the parent article. The single beam spectra of the sample and reference were ratioed to obtain both log(1/R) and Kubelka-Munk spectra separately using the Thermo-Nicolet OMNIC software version 7.0.

The MSC algorithm was applied to log(1/R) spectra first. The baseline offset for each spectrum was obtained by averaging from 4840 to 5000 cm^{-1} for caffeine and from 4670 to 4840 for EMPA. Then each baseline offset was subtracted from its spectrum. Finally those log(1/R) spectra were converted into Kubelka-Munk. It is important to apply MSC first and then baseline offset subtraction to the log(1/R) spectrum. This is because MSC is not a baseline correction algorithm, and if the order is reversed the

‘minor’ baseline offset it leaves out will further contribute to the baseline position and to the peak height in the converted Kubelka-Munk spectra as described before.

The KMC algorithm was applied separately to the original MSC corrected log(1/R) spectra followed by conversion to KM units, and to the Kubelka-Munk MSC corrected spectra. The standard deviation as a function of wavenumber was calculated on those multiple measurements of the same sample for both original log(1/R) and converted Kubelka-Munk spectra (see Figures S-3 and S-4). Close examination reveals that converted Kubelka-Munk spectra have slightly higher spectral intensities relative to those of the original log(1/R) spectra. The reason for this difference is unknown.

Comparison of original and converted Kubelka-Munk spectra and their standard deviations for both the caffeine and the EMPA data demonstrates that there is no significant difference between those two kinds of spectra.

Conclusion

Both theoretical and experimental studies have proved that our KMC correction in Kubelka-Munk space is equivalent to applying MSC first, then subtracting the baseline offset in log(1/R) space. Our KMC correction has some practical advantages over the corrections in log(1/R) by virtue of (1) the existence of a linear relationship between Kubelka-Munk and sample concentration and (2) the ease with which MSC and BPC (Eq. 4) can be implemented since they are both matrix operations applied to every wavenumber component over the entire spectral range.

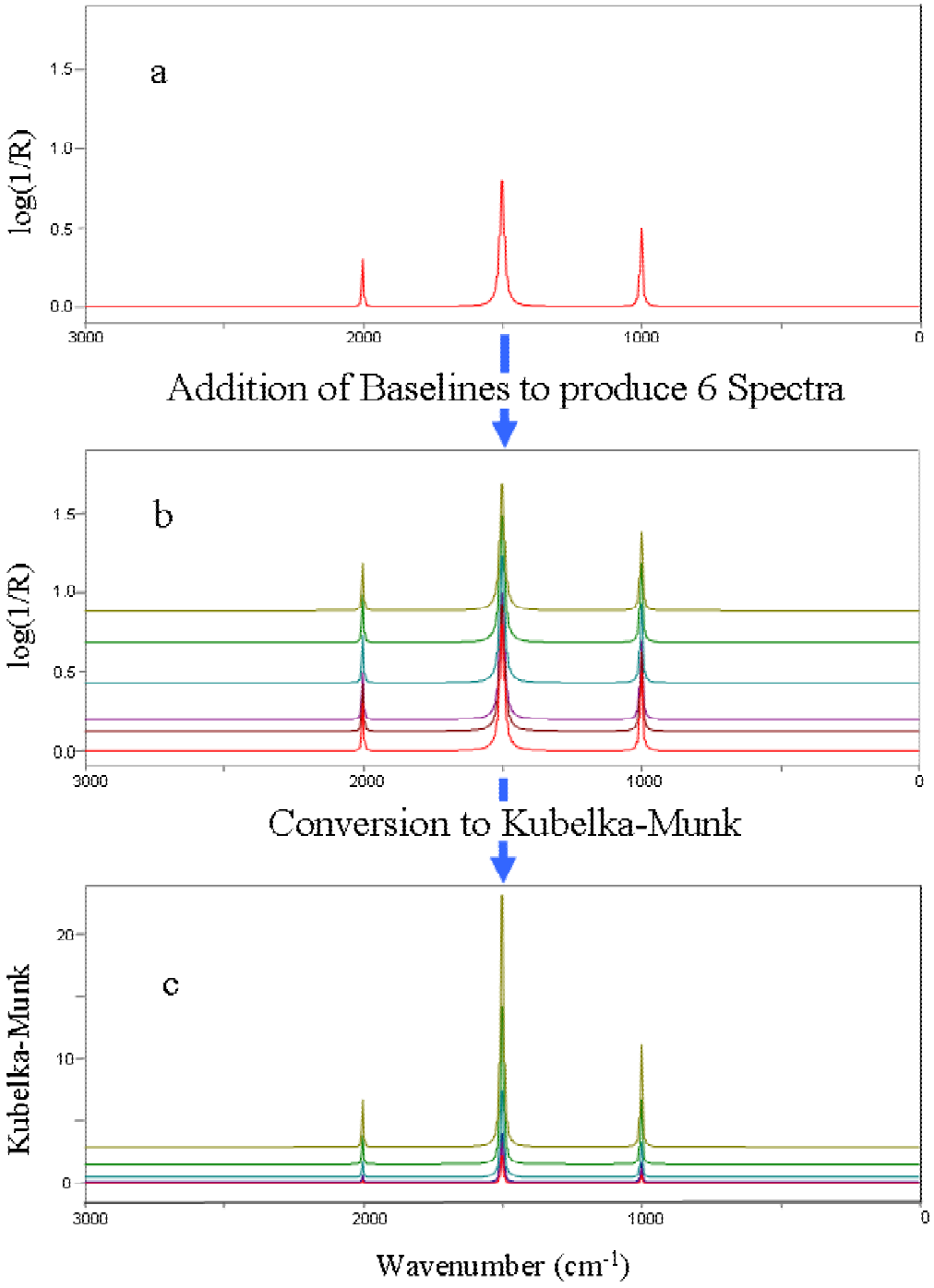
Supporting information figure captions:

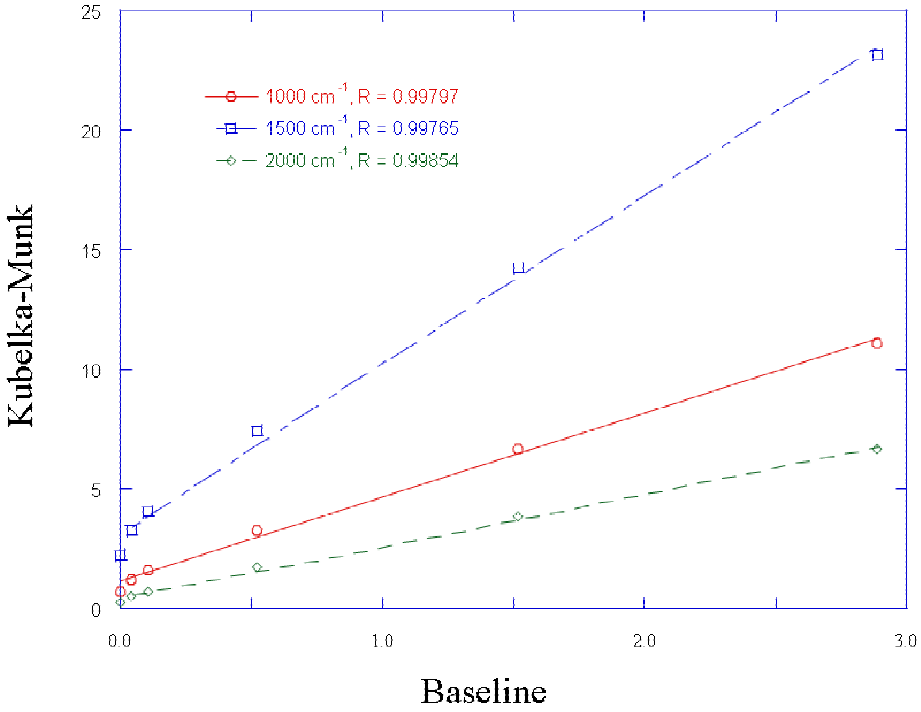
Figure S-1 Synthesized spectrum study with baseline addition and conversion from log(1/R) into Kubelka-Munk units. Figure S-1a shows three bands located at 1000, 1500 and 2000 cm^{-1} from right to left. Six baseline offsets added result in six new reflection spectra in Figure S-1b. Conversion from log(1/R) into Kubelka-Munk produces the spectra in Figure S-1c.

Figure S-2 Peak intensity vs. baseline offset plot for all three synthesized bands in Kubelka-Munk space. The least square linear fitting results of three bands are displayed in the legend of the plot.

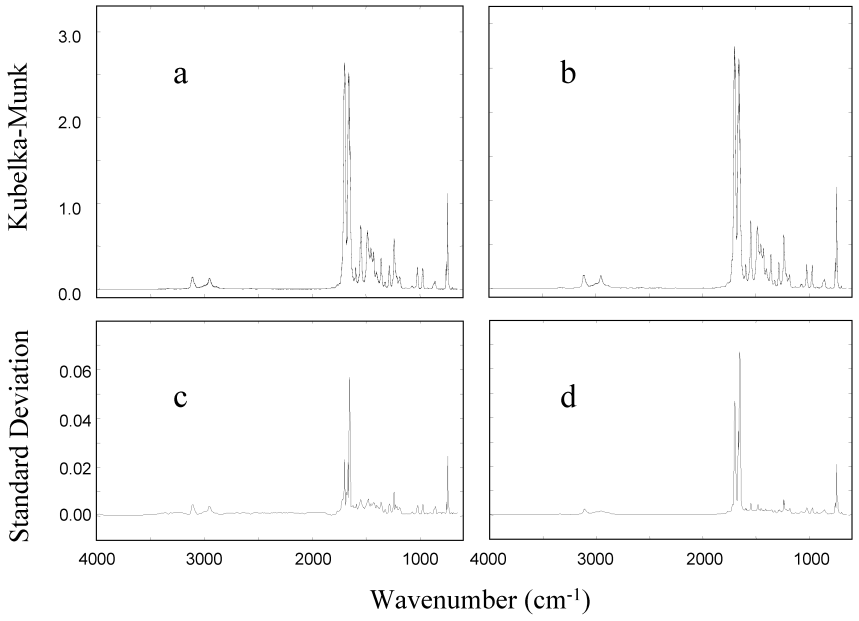
Figure S-3 The original and converted caffeine Kubelka-Munk spectra after applying suitable corrections in Kubelka-Munk and log(1/R) spaces separately. Figure S-3a is the KMC-corrected data in log(1/R) space converted into Kubelka-Munk units. Figure S-3b is the KMC-corrected data in Kubelka-Munk space. Figure S-3c and d are the standard deviation spectra corresponding to Figure S-3a and b.

Figure S-4 The overlaid and standard deviation of original and converted EMPA/soil Kubelka-Munk spectra after applying suitable corrections in Kubelka-Munk and log(1/R) spaces separately. Figure S-4a is the KMC-corrected data in log(1/R) space converted into Kubelka-Munk units. Figure S-4b is the KMC-corrected data in Kubelka-Munk space. Figure S-4c and d are the standard deviation spectra corresponding to Figure S-4a and b.

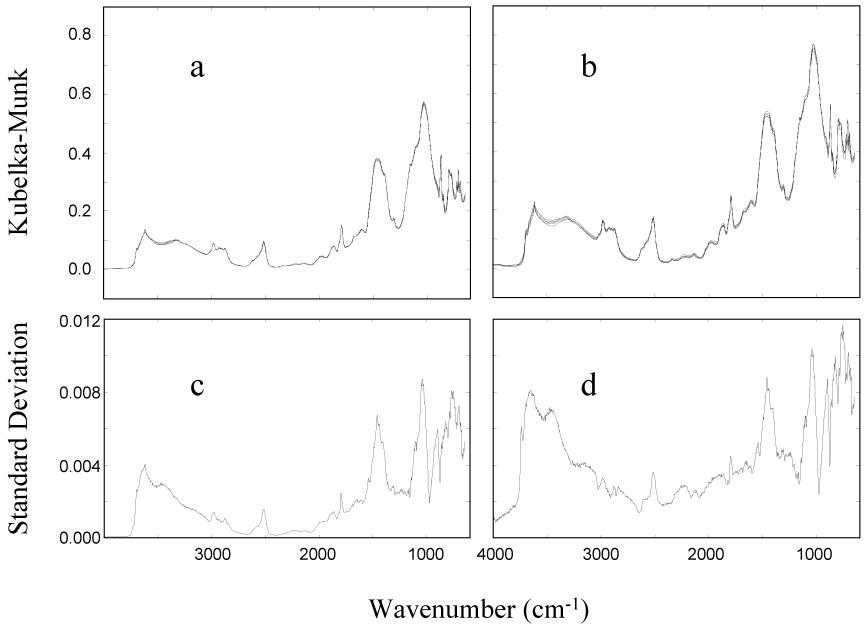

a
log(1/R)
1.5
1.0
0.5
0.0
3000
2000
1000
0
Addition of Baselines to produce 6 Spectra
b
log(1/R)
Conversion to Kubelka-Munk
c
Kubelka-Munk
20
10
0
Wavenumber (cm-1)


1000 cm-1, R = 0.99797
1500 cm-1, R = 0.99765
2000 cm-1, R = 0.99854
Kubelka-Munk
Baseline
0
5
10
15
20
25
0.0
1.0
2.0
3.0

S-2


a
b
c
d
Kubelka-Munk
Standard Deviation
Wavenumber (cm-1)
0.0
1.0
2.0
3.0
0.00
0.02
0.04
0.06
4000
3000
2000
1000

S-3


Kubelka-Munk
Standard Deviation
a
b
c
d
0.8
0.6
0.4
0.2
0.0
0.012
0.008
0.004
0.000
3000
2000
1000
4000
3000
2000
1000
Wavenumber (cm-1)