

Introduction to USGSHydroOpt

Samuel Christel¹ and Steve Corsi¹

¹*United States Geological Survey*

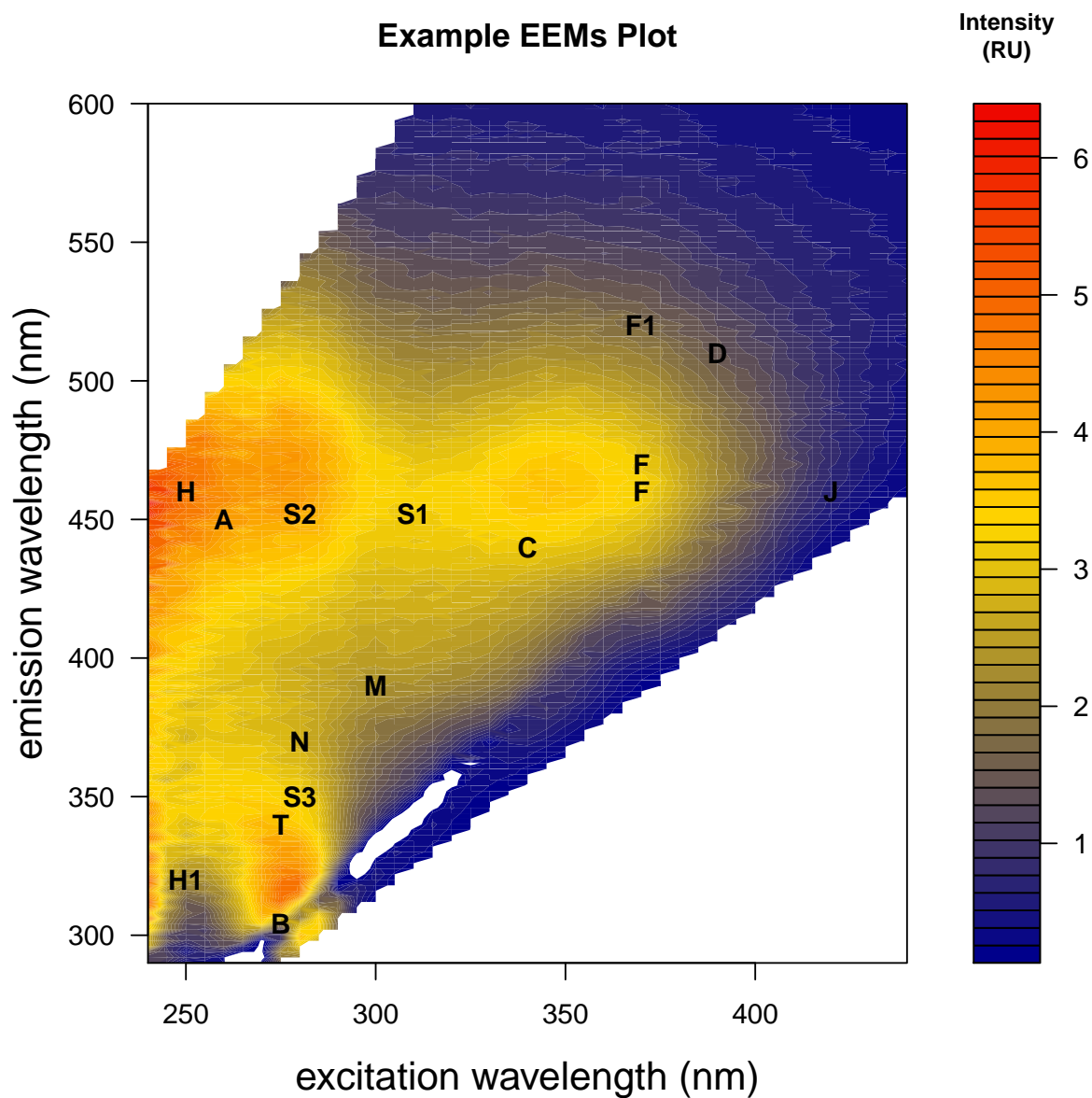
June 24, 2014

Contents

1	Introduction to USGSHydroOpt	1
2	Dataframe Formatting for USGS HydroOpt	2
2.1	Absorbance Data	2
2.2	Fluorescence Data	3
2.3	Spectral Slopes Data	3
2.4	Optical Summary Data	4
2.5	Excitation-Emission (EEMs) Peak Data	4
2.6	Optical Ratio and Signals Data	5
2.7	Optical Signals Data	6

1 Introduction to USGSHydroOpt

The USGSHydroOpt package was created to streamline the process of creating optical summary variables and excitation-emission (EEMs) plots for absorbance and fluorescence data collected from various freshwater sources. Examples of optical summary variables that can be produced with this package include various absorbance peaks, The functions in this package were designed to operate on dataframes with a standardized structures. This package is not amenable to dataframes that do not fit the prescribed format. The example dataframes in this package illustrate exactly how dataframes should be formatted, and the examples illustrate how the functions operate on the dataframes. Depicted below is an example of an EEMs plot produced with this package.



NULL

2 Dataframe Formatting for USGS HydroOpt

The functions contained in USGSHydroOpt operate on dataframes with defined structures. Users interested in using USGSHydroOpt should format dataframes according to the structures defined in this section.

2.1 Absorbance Data

Absorbance data used by functions in USGSHydroOpt should be formatted such that each sample occupies a column, and one column contains the wavelength (nm) for which the absorbance measurement was mea-

sured (*See example Table 2.1 below*). The column with the wavelengths in Table 2.1 does not need to be called "wavelengths," as it is named in the example dataframe below. Since this package was developed primarily for USGS activities, the default for naming samples is "gr" then the sample number. This convention was started by the USGS California Water Sciences Center (CA WSC) and the USGS Wisconsin Water Science Center (WI WSC) follows the same naming convention to ensure standardization.

	gr13307	gr13351	gr13353	gr13357	wavelengths
1	0.0001296	-0.0003505	-0.0003480	-0.0002695	750
2	-0.0002367	0.0000305	-0.0000915	-0.0001407	749
3	-0.0001582	-0.0004900	-0.0006325	-0.0001534	748
4	-0.0004642	-0.0000105	-0.0000932	-0.0000245	747
5	-0.0002551	-0.0000653	0.0000841	-0.0001615	746
6	-0.0001842	-0.0002135	0.0002082	0.0001429	745

2.2 Fluorescence Data

Fluorescence data used by functions in USGSHydroOpt should also be formatted such that each sample occupies a column, and one column contains the excitation emission wavelength pairs (nm) for which the fluorescence measurement was measured (*See example Table 2.2 below*). The column with the excitation emission pairs in Table 2.2 does not need to be called "Wavelength.Pairs," as it is in the example dataframe below. Again since this package was developed for USGS activities, the default sample naming convention is "gr" followed by the sample number.

	Wavelength.Pairs	gr13307	gr13308	gr13351	gr13352
1	240/290	0.09457	0.05358	-0.07268	-0.19368
2	240/292	0.07386	0.09194	0.45433	0.39722
3	240/294	0.08116	0.09930	0.49135	0.37943
4	240/296	0.12013	0.09138	-0.12248	0.09239
5	240/298	0.15264	0.15428	0.54694	0.14980
6	240/300	0.16467	0.13010	0.43319	0.33064

2.3 Spectral Slopes Data

Information on the upper and lower wavelength (nm) for which a spectral slope should be calculated needs to be stored in a dataframe if USGSHydroOpt is used. The dataframe should contain exactly three columns. The first column should contain the upper wavelength, the second column should contain the lower wavelength, and the third column should contain the name of the spectral slope being calculated (*See example Table 2.3 below*). The columns in Table 2.3 need to be in this exact order, although the names of the columns may be different. The data types for each column are integer, integer, and character, respectively. More spectral slopes can be added to the table than specified in the example dataframe below.

	wvlngh1	wvlngh2	Name
1	275	290	Sag275_295
2	290	350	Sag290_350
3	350	400	Sag350_400
4	412	676	Sag412_676

2.4 Optical Summary Data

This is the dataframe that contains many of the summary optical variables that can be produced using functions in USGSHydroOpt (*See example Table 2.4a below*). The functions in USGSHydroOpt calculate summary optical variables and add to a dataframe formatted according to Table 2.4a below. The example dataframe below is how the WI WSC stores optical summary variables. Note that this dataframe can contain other columns with metadata, for example, the sample data and time, the sample ID, or whether or not the sample went through QA/QC.

	GRnumber	B	T	A	J	FI_2005	A254
1	gr13307	0.050997	0.1826	0.4553	0.03705	1.639	0.05228
2	gr13351	-0.245602	0.4085	3.0783	0.19110	1.456	0.43531
3	gr13353	-0.175220	0.7794	6.8624	0.56309	1.523	0.68274
4	gr13357	0.111561	0.2691	0.9451	0.06969	1.572	0.08698
5	gr13360	-0.001569	0.4593	2.8254	0.37231	1.563	0.28605
6	gr13363	0.052137	0.4892	1.9518	0.19098	1.583	0.19283

However, also note that summary optical variable names in the dataframe must be identical to those specified in Table 2.4b below.

[1]	"OB1"	"OB2"	"OB3"	"S1.50"
[5]	"S2.50"	"S3.50"	"S1.25"	"S2.25"
[9]	"S3.25"	"Mrange.25"	"Mrange.50"	"B"
[13]	"T"	"M"	"A"	"C"
[17]	"N"	"D"	"F"	"J"
[21]	"S1"	"S2"	"S3"	"H1"
[25]	"H2"	"F1"	"F2"	"W"
[29]	"LT1"	"LT2"	"LT3"	"LA"
[33]	"HIX_2002"	"FI_2005"	"FI_2001"	"FreshI"
[37]	"A254"	"A275"	"A280"	"A290"
[41]	"A295"	"A350"	"A370"	"A400"
[45]	"A412"	"A440"	"A488"	"A510"
[49]	"A532"	"A555"	"A650"	"A676"
[53]	"A715"	"Sag275_290"	"Sag290_350"	"Sag350_400"
[57]	"Sag412_676"	"Aresids"		

2.5 Excitation-Emission (EEMs) Peak Data

The EEMs peak data contains three columns listing the name of a characterized EEM peak along with the corresponding wavelengths (nm) (*See example Table 2.5 below*). The first column, in Table 2.5, "Peak," contains the name of the characterized EEM peak. The next two columns in Table 2.5 contain the excitation and emission wavelengths (nm) at which a given peak occurs. The column names must be identical to those displayed in the example below, although the order of the columns can be different.

	Peak	ExCA	EmCA
1	B	275	304
2	T	275	340
3	M	300	390
4	A	260	450

5	C	340	440
6	N	280	370
7	D	390	510
8	F	370	460
9	J	420	460
10	S1	310	452
11	S2	280	452
12	S3	280	350
13	H	250	460
14	H1	250	320
15	F	370	470
16	F1	370	520

2.6 Optical Ratio and Signals Data

This dataframe contains one column called "ratioSignals" that contains all of the summary optical variables currently identified by the WI WSC (*See example Table 2.6 below*). Note that these are the same variables as those listed in Table 2.4b. The first column must contain the various "ratioSignals" that the user desires, although the column name need not be "ratioSignals"

	ratioSignals	keep
1	OB1	NA
2	OB2	NA
3	OB3	NA
4	S1.50	NA
5	S2.50	NA
6	S3.50	NA
7	S1.25	NA
8	S2.25	NA
9	S3.25	NA
10	Mrange.25	NA
11	Mrange.50	NA
12	B	NA
13	T	NA
14	M	NA
15	A	NA
16	C	NA
17	N	NA
18	D	NA
19	F	NA
20	J	NA
21	S1	NA
22	S2	NA
23	S3	NA
24	H1	NA
25	H2	NA
26	F1	NA
27	F2	NA

28	W	NA
29	LT1	NA
30	LT2	NA
31	LT3	NA
32	LA	NA
33	HIX_2002	NA
34	FI_2005	NA
35	FI_2001	NA
36	FreshI	NA
37	A254	1
38	A275	1
39	A280	1
40	A290	1
41	A295	1
42	A350	1
43	A370	1
44	A400	1
45	A412	NA
46	A440	NA
47	A488	NA
48	A510	NA
49	A532	NA
50	A555	NA
51	A650	NA
52	A676	NA
53	A715	NA
54	Sag275_290	1
55	Sag290_350	1
56	Sag350_400	1
57	Sag412_676	1
58	Aresids	NA

2.7 Optical Signals Data

This dataframe is similar to "ratioSignals" except it provides more metadata about peaks characterized for EEMs plots. The dataframe should contain six columns (*See example Table 2.7 below*). The first column in Table 2.7, "Peak," contains the name of the characterized EEM peak. The next two columns, "Ex1" and "Ex2," contain the excitation wavelength range (nm) for a given peak. "Ex1" is the lower wavelength and "Ex2" is the upper wavelength for the excitation wavelength range (nm) for a given peak. Similarly, "Em1" and "Em2" contain the emission wavelength range (nm) for a given peak. The final column, "Source," lists the source that characterized the peak. The last column, "Source," is not required. The user must exactly replicate the column names in Table 2.7 in order for the code in USGSHydroOpt to run.

	Peak	Ex1	Ex2	Em1	Em2	Source
1	OB1	360	NA	410	598	Hartel Turner
2	OB2	360	NA	436	436	Hartel Turner
3	OB3	365	NA	400	550	Hagedorn Turner
4	S1.50	310	NA	402	502	Sniffer

5	S2.50	280	NA	402	502	Sniffer
6	S3.50	280	NA	310	390	Sniffer
7	S1.25	310	NA	427	477	Sniffer
8	S2.25	280	NA	427	477	Sniffer
9	S3.25	280	NA	330	370	Sniffer
10	Mrange.25	300	NA	365	415	test
11	Mrange.50	300	NA	340	440	test
12	B	275	NA	304	NA	CA
13	T	275	NA	340	NA	CA
14	M	300	NA	390	NA	CA
15	A	260	NA	450	NA	CA
16	C	340	NA	440	NA	CA
17	N	280	NA	370	NA	CA
18	D	390	NA	510	NA	CA
19	F	370	NA	460	NA	CA
20	J	420	NA	460	NA	CA
21	S1	310	NA	452	NA	CA
22	S2	280	NA	452	NA	CA
23	S3	280	NA	350	NA	CA
24	H1	250	NA	460	NA	Ohno2002
25	H2	250	NA	320	NA	Ohno2002
26	F1	370	NA	470	NA	Cory and McKnight, 2005
27	F2	370	NA	520	NA	Cory and McKnight, 2005
28	W	255	290	302	350	
29	LT1	250	NA	340	NA	
30	LT2	260	NA	340	NA	
31	LT3	240	NA	340	NA	
32	LA	240	NA	440	NA	