
Spike Detection And Correction

Introduction

All MKIV Brewers in the NEUBrew network produce relatively frequent spikes during UV scans. The spikes are often very large. If not corrected a significant error affecting UV index calculations results. The origin of the spikes is not really known. We certainly do not think that they are caused by cosmic ray events. We agree with Meinander et al. (2003) that the most likely origin of the spikes is some glitch in the Brewer's hardware, firmware and/or software. It is interesting that newer MKIII Brewers (after year 2000) supposedly do not exhibit this behavior.

The difficulty in detecting the signal anomaly is due to large amplitude and frequency of Fraunhofer lines structure imposed on the signal and due to UV signal variation with changing sky conditions during a scan. The latter would not pose a difficulty in detector array based instruments that take instantaneous snapshots of the whole spectrum. But in a scanning instrument like the Brewer each wavelength measurement is taken under different sky conditions. When sky conditions are changing rapidly, particularly near sun azimuth and zenith, real changes in total horizontal irradiance may manifest itself like signal anomalies.

Any algorithm used in detecting signal anomalies for a given class of signals has its inherent probability of missed detection (PMD) and probability of false detection (PFD). Since the ranges of values of signals with and without spikes overlap, it is practically impossible to detect all spikes and avoid making false detections. Nevertheless it is possible to devise a procedure that makes both PMD and PFD as small as possible. Spikes in the Brewer seem to have large amplitudes - often significantly larger than the signal. These spikes can be detected and corrected almost 100% of the time. On the other hand frequent signal anomalies of low amplitude with spike attributes can be ignored without great consequence to signal fidelity.

Spike detection

We follow Meinander et al. (2003) in our approach to detect spikes. First, at wavelength λ_i we calculate the ratio $r_i = s_i / \text{ref}_i$, where s_i is the normalized signal, i.e., $s_i = S_i / \sum S_j$, where $\sum S_j$ is the signal sum over all 154 wavelengths. And ref_i is a value of a reference scan also normalized by its sum ($\text{ref}_i = \text{REF}_i / \sum \text{REF}_j$). The reference signal is obtained from several clear sky and spike free scans close to a solar noon (Fig.1). The spike detection scheme uses differences of the ratio: $\Delta r_i = r_i - r_{i-1}$.

First the mean $\mu(\Delta r_i)$ and standard deviation $\sigma(\Delta r_i)$ of the large set of data (we have used up to 15,000 scans) of Δr_i are estimated. The detection criteria are defined as follows: if

$$\Delta r_i - \mu(\Delta r_i) > k\sigma(\Delta r_i) \quad \& \quad \Delta r_{i+1} - \mu(\Delta r_{i+1}) < -k\sigma(\Delta r_{i+1}) \quad (1a)$$

a positive spike is detected and if

$$\Delta r_i - \mu(\Delta r_i) < -k\sigma(\Delta r_i) \quad \& \quad \Delta r_{i+1} - \mu(\Delta r_{i+1}) > k\sigma(\Delta r_{i+1}) \quad (1b)$$

a negative spike is detected. For $k=3$ over 99% of suspect spikes should be found. This is not necessarily true as the statistics of spike sizes is not necessarily Gaussian. However in general for $k=3$ we may expect to find one spike suspect in every scan as the criteria (1a) and (1b) are tested 152 times in one scan.

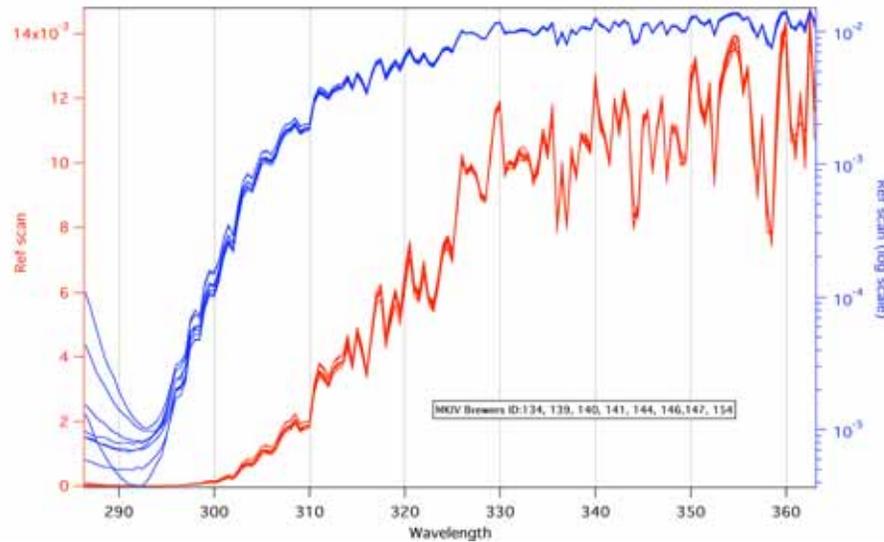


Figure 1. Normalized reference scans used in the spike detection scheme.

We have added five modifications (M1,...,M5) to the original Meinander's criteria.

(M1) If the standard deviation $\sigma_P(\Delta r_i)$ of the Poisson photon noise exceeds the standard deviation $\sigma(\Delta r_i)$, then in (1a) and (1b) $\sigma_P(\Delta r_i)$ is used instead of $\sigma(\Delta r_i)$. This often happens when the signal is small as is usual in early morning or late afternoon. The standard deviation $\sigma_P(\Delta r_i)$ is calculated as follows:

$$\sigma_P(\Delta r_i) = [(\Delta s_i / \text{ref}_i)^2 + (\Delta s_{i-1} / \text{ref}_{i-1})^2]^{1/2} \quad (2)$$

where Δs_i is the standard deviation of the Poisson noise in the signal s_i at wavelength λ_i .

(M2) When either (1a) or (1b) is triggered, the ratio at r_i is recalculated on the fly as i is incremented:

$$r_i^c = 0.5 (r_{i-1} + r_{i+1}) \quad (3)$$

and values $\Delta r_i = r_i^c - r_{i-1}$ and new $\Delta r_{i+1} = r_{i+1} - r_i^c$ are updated. This update of Δr_i eliminates cases when two same sign spikes at λ_i and λ_{i+2} may cause a false detection of spike of opposite sign in between at λ_{i+1} .

(M3) Sometimes a large spike at λ_i triggers (1a) or (1b) to detect a false small spike at λ_{i-1} or λ_{i+1} . This can be avoided by checking conditions (1a) and (1b) twice. First the algorithm is run from left to right (λ_i increasing) and then a second time from right to left (λ_i decreasing). Then the detections that are not coincident nor of the same sign in both runs are discarded.

(M4) The conditions (1a) and (1b) cannot be applied for the first ($i=0$) and the last ($i=153$) wavelengths. Spikes at $\lambda_0=286.5\text{nm}$ can be ignored as signal there is hidden in the stray light and subsequently discarded, but spikes at $\lambda_{153}=363\text{nm}$ are more important. So for λ_{153} we detect a spike as follows:

$$|r_{153} - r_{153}^c| / r_{153}^c > t_{153} \quad (4)$$

where $r_{153}^c = 0.5(r_{151} + r_{152})$ and t_{153} is the detection threshold.

(M5) It is fortunate that for most of the Brewers that we have tested negative spikes are less frequent than positive spikes. A negative spike is likely to be mistaken with a signal drop caused by a passage of a small fast moving cloud that obscures the sun when a UV scan is performed. It is unlikely that the signal drop applies to one wavelength only as the time of measurement at one wavelength is only ≈ 1.7 seconds. Conditions (1a) or (1b) will not be triggered in cases of two consecutive signal drops of similar size. However, three consecutive signal drops at λ_i , λ_{i+1} and λ_{i+2} , where at λ_{i+1} the signal drop is smaller than at the two neighboring wavelengths, can be detected as two spikes at λ_i and λ_{i+2} . We check the relative values of the two spikes by calculating the following:

$$a = 0.5 \cdot (r_i^c + r_{i+2}^c) - 0.5 \cdot (r_i + r_{i+2}) \quad \text{and} \quad b = r_{i+1}^c - 0.5 \cdot (r_i + r_{i+2})$$

The double-spike detection is cancelled if

$$|b/a| < t_c \quad (5)$$

implying that it was a cloud passage case.

Spike correction

Not all detected spikes with criteria (1a) or (1b) are corrected. We setup two thresholds $R_{\text{corrected}}$ and R_{flagged} , which are applied to spikes of relative magnitude $M_i = r_i / r_i^c - 1$ If

$$|r_i / r_i^c - 1| > R_{\text{corrected}} \quad (6a)$$

then the spike is corrected; if

$$R_{\text{corrected}} \geq |r_i/r_i^c - 1| > R_{\text{flagged}} \quad (6b)$$

then the spike is flagged; and if

$$R_{\text{flagged}} \geq |r_i/r_i^c - 1| \quad (6c)$$

then the spike is ignored.

The corrected signal in case (6a) is calculated as follows:

$$s_{i=}^c = r_i^c \cdot \text{ref}_i \quad \text{and} \quad S_{i=}^c = s_{i=}^c \sum S_j \quad (7)$$

The scaling with the reference signal ref_i in (7) retains the Fraunhofer lines signature in the corrected value with respect to the neighboring wavelengths.

Values of Parameters

The choice of parameters is to some extent arbitrary. Our definition of a flagged only spike is arbitrary. Some parameters are complementary and interact with each other. For instance when k is small the number of detected spikes increases but since “real” spikes seem to have large amplitudes it does not affect the number of corrected spikes that much. Within a limit the number of corrected spikes is not that sensitive to the $R_{\text{corrected}}$ value either. Still, as is shown below, different instruments display different spike behavior. The number of detections and corrections are different for different instruments. The number of spikes is a function of instrument (large spikes) and location (flagged spikes). The latter might be due to different prevailing sky conditions. Certainly, more cases of clear or overcast skies leads to more robust detections of even small spikes and more cases of partly cloudy skies may lead to more false detections of small spikes.

It might be possible to fine tune the parameters for each instrument separately. However we have run the procedure with the same parameters on eight MKIV Brewers having some evidence that further fine tuning of parameters does not make that much difference.

The following parameters were used: $k=3$, $t_{153}=0.25$, $t_c=0.65$, $R_{\text{corrected}}=0.5$, $R_{\text{flagged}}=0.15$

We keep in mind that changes in instrument responsivity would call for a different reference scan and even (less likely) update of mean $\mu(\Delta r_i)$ and standard deviation $\sigma(\Delta r_i)$. For this reason the statistics of spike corrections and detections have to be monitored. Obviously responsivity altering changes like replacement of optical elements (diffuser, filter) or replacement of the PMT most likely would require a new reference scan. When an instrument is relocated to a new site with different climatology a new mean $\mu(\Delta r_i)$ and standard deviation $\sigma(\Delta r_i)$ should be acquired. On the other hand changes in parameters should not be required when responsivity or location is changed.

Bad scan detection

We use three criteria to flag bad scans.

(a) Signal is not strong enough if

$$\langle r \rangle_{[286.5,294.0]} / \langle r \rangle_{[325.5,363.0]} > \varepsilon \quad (8a)$$

where $\langle r \rangle_{[286.5,294.0]}$ is the average of ratio r_i values in [286.5,294.0] region. The threshold $\varepsilon=3.0$ was used successfully with data from all Brewers tested.

(b) Signal has too large of a jump at $i=77$ (325nm) if

$$|(\langle r \rangle_{[323.5,325.0]} - \langle r \rangle_{[325.5,327.0]}) / \langle r \rangle_{[323.5,327.0]} - \mu_{77}| > \varepsilon_{77} \quad (8b)$$

where μ_{77} is the expected value obtained from the instrument data statistics and the threshold $\varepsilon_{77}=0.55$ seems to catch all anomalies for all tested instruments without false detections.

(c) Spike and stray light corrected signal's cut-on wavelength is too long. The cut-on wavelength λ_{cuton} is defined as the longest wavelength at which the signal is non-positive. If $\lambda_{\text{cuton}} > 321\text{nm}$ the scan is flagged as bad.

The criterion (a) is applied before (b). It seems that criterion (a) catches most of the scans that otherwise would be caught by (b). This suggests that bad scans of type (a) and (b) have similar origins in instrument malfunction. We speculate that (a) and (b) might be caused by one of the filters covering the beam aperture only partially or when a wrong filter is in the beam aperture during a UV scan.

Results

The statistical results from the data run on all eight Brewers is collected in Table 1. The same set of detection parameters were used in all cases.

Without knowing the origin of the spikes it is hard to explain the differences between different Brewers. BR146 has the lowest spike rate (50 scans/1spike) whereas BR139 and BR140 have the highest rate (7scans/1spike) while the values for the remaining five Brewers are (13-18 scans/1spike). Except for BR146 all those rates are much higher than the MKII BR037 tested by the Finnish Meteorological Institute (Meinander et al. 2003)

The fraction of negative spikes is different among Brewers. In all cases all flux in $\text{mW/m}^2/\text{nm}$ of corrected spikes is 100-3000 times higher than in the flagged spikes. In

fact, for all cases the flagged spikes on average contain less than 0.33 mW/m²/nm irradiance.

Table 1. Results for eight MKIV Brewers

Brewer ID#	Number of Tested Scans	Number of Corrected Spikes	Corrected Spike Rate [Scans/1spike]	Number of Negative Spikes	Average Spike Flux (Corrected) [mW/m ² /nm]	Number of Flagged Spikes	Flagged Spike Rate [Scans/1spike]	Average Spike Flux (Flagged) [mW/m ² /nm]	Number of Bad Scans
134	14599	931	15	12	315.3	1428	10	0.14	9
139	7417	1061	7	408	365.1	617	12	0.10	23
140	15238	907	17	160	316.2	1670	9	0.14	5
141	13234	1577	8	11	397.3	1070	12	0.08	13
144	11519	677	17	6	236.4	1195	10	0.18	11
146	13465	270	50	2	255.6	4757	3	0.15	14
147	12907	988	13	5	324.1	8821	1.5	0.18	9
154	15118	856	18	14	316.3	2466	6	0.33	4

A significantly higher rate (4-8 times) of flagged spikes occurs for BR146 located at the Mountain Research Station, Colorado and for BR147 located at Fort Peck, Montana. We are not certain whether the high rate of small spikes in these cases is a feature of the instruments (two consecutive BR numbers) or the prevailing dynamic sky conditions.

The highest number of corrected negative spikes occurs for BR139 and BR140 located at Table Mountain, Colorado and Raleigh, North Carolina, respectively. We do not know why these two instruments have a distinct behavior. BR139 is collocated with BR134 and BR141, both of which have very low rates of negative spikes. It might not be a coincidence that the two instruments have consecutive serial numbers.

In the Appendix detailed statistical results for each instrument are presented in three separate panels in the form of tables and graphs.

References

Meinander, O., W. Josefsson, J. Kaurola, T. Koskela, and K. Lakkala, "Spike detection and correction in Brewer spectroradiometer ultraviolet spectra.," *Opt. Eng.* **42**, 1812-1819, 2003.

Appendix: Results for individual instruments

Panel: PARAMETERS

Graph 1 (1st from the top) shows values of the mean $\mu(\Delta r_i)$ and the standard deviation $\sigma(\Delta r_i)$ obtained from all data (dotted red lines) and after the spikes and bad scans were removed (blue lines). The latter are used in the detection scheme.

Graph 2 (2nd from the top) shows the number of spikes above the $R_{corrected}$ threshold as a function of the threshold for negative spikes (blue) and positive spikes (red). The plateau around $R_{corrected} = 0.5$ indicates the setting for the optimal $R_{corrected}$ threshold.

Graph 3 (3rd from the top) shows the spike rate (scans/spike) for six (A, ..., F) runs with different parameters. The parameters are listed in the first column of the table that follows. In Case A, dotted red (from Graph 1) the mean $\mu(\Delta r_i)$ and the standard deviation $\sigma(\Delta r_i)$ are used. In Case B, continuous blue (from Graph 1) the mean $\mu(\Delta r_i)$ and the standard deviation $\sigma(\Delta r_i)$ are used. This results in a larger number of detections. In Case C the criterion M1 is introduced resulting in lower spike detection presumably because of a reduction of false detections. In Cases C, D, E and F the threshold is progressively increased from 0.15 to 0.5, which results in lowering the number of corrected spikes until the plateau is reached.

The table also list the number of spike suspects that were cleared by the criterion M3 as N_{coinc} . This number is significant. On the other hand N_{cld} the number of double-spike suspects that were cleared by the criterion M5 is small.

Panel: SPIKES STATISTICS 1

Graph 1 (1st from top) shows the number of spikes per scan (blue) and their total irradiance (red). For some instruments the Graph 1 hints on seasonal dependence.

Graph 2 (2nd from top) shows a histogram of spike irradiance of all corrected spikes.

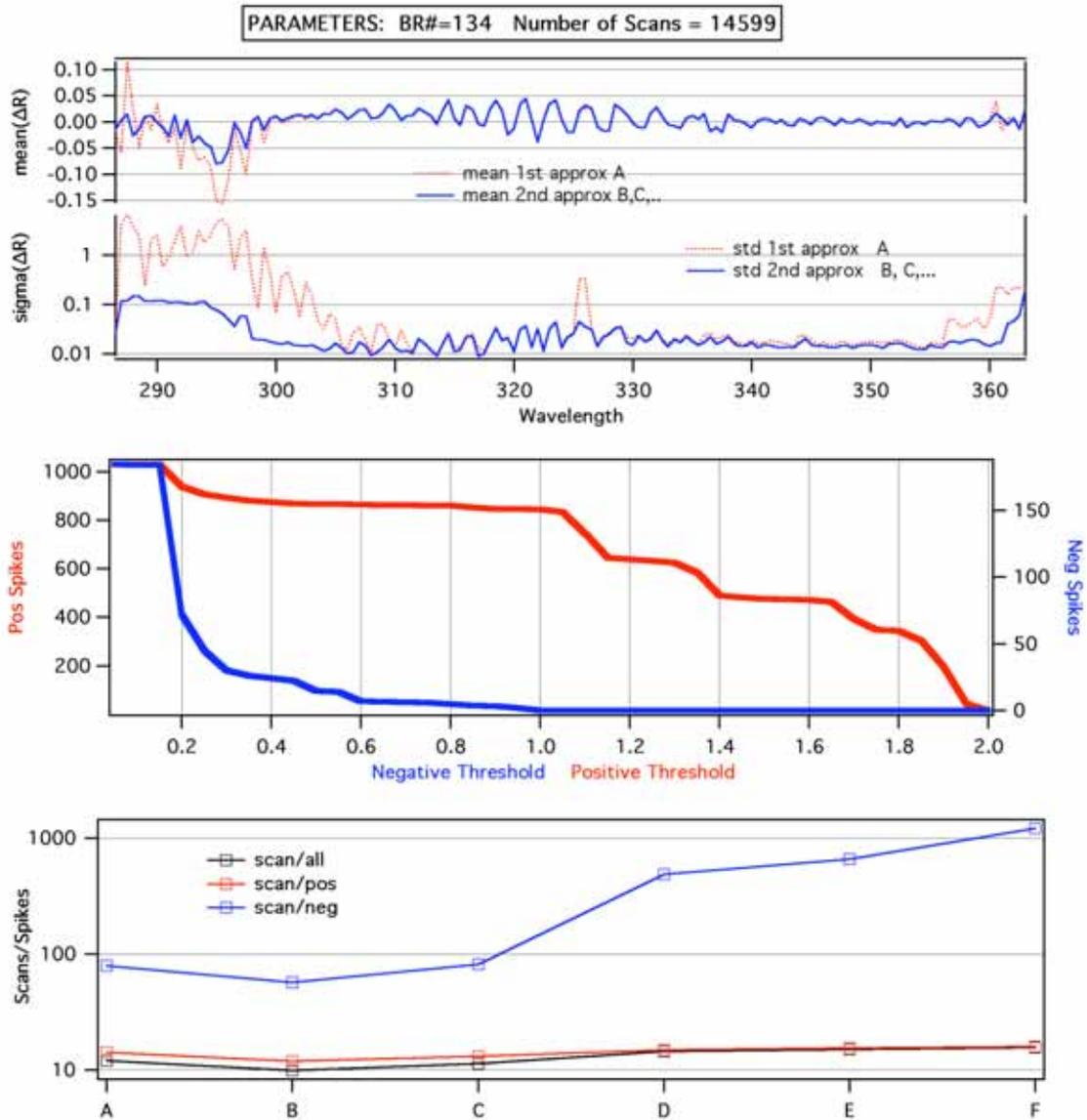
Graph 3 (3rd from top) shows spike distribution as a function of wavelength. The distribution is not random and it differs from instrument to instrument.

Panel: SPIKES STATISTICS 2

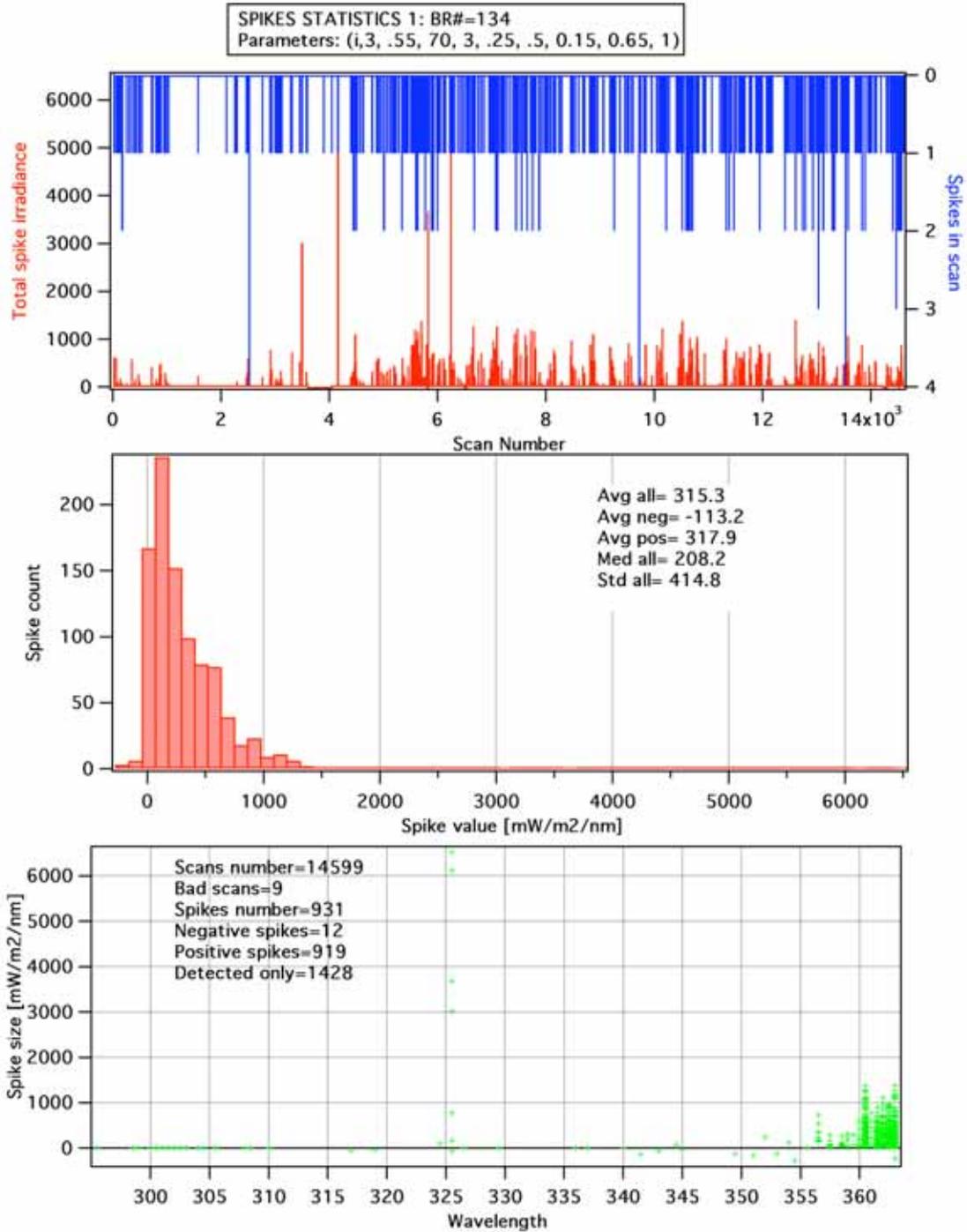
Graph 1 (1st from top) shows distribution of corrected (red) and flagged (blue) spikes with respect of wavelength. In most case flagged (blue) show more random distribution.

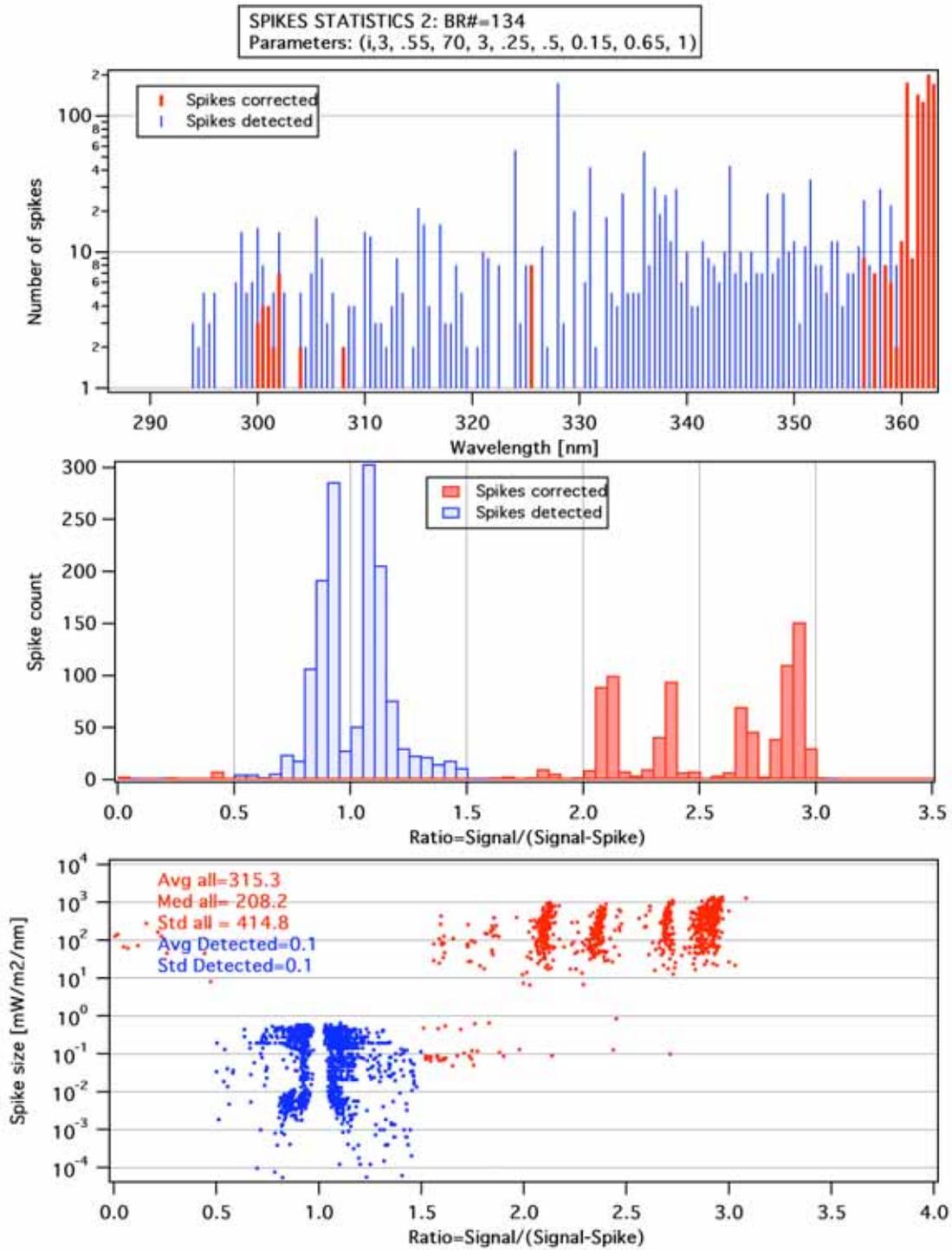
Graph 2 (2nd from top) shows two histograms of spike frequency with respect to the spike ratio $M_{i+1} = r_i / r_i^c$. One can easily see the distinction between corrected (red) and flagged (blue) spikes.

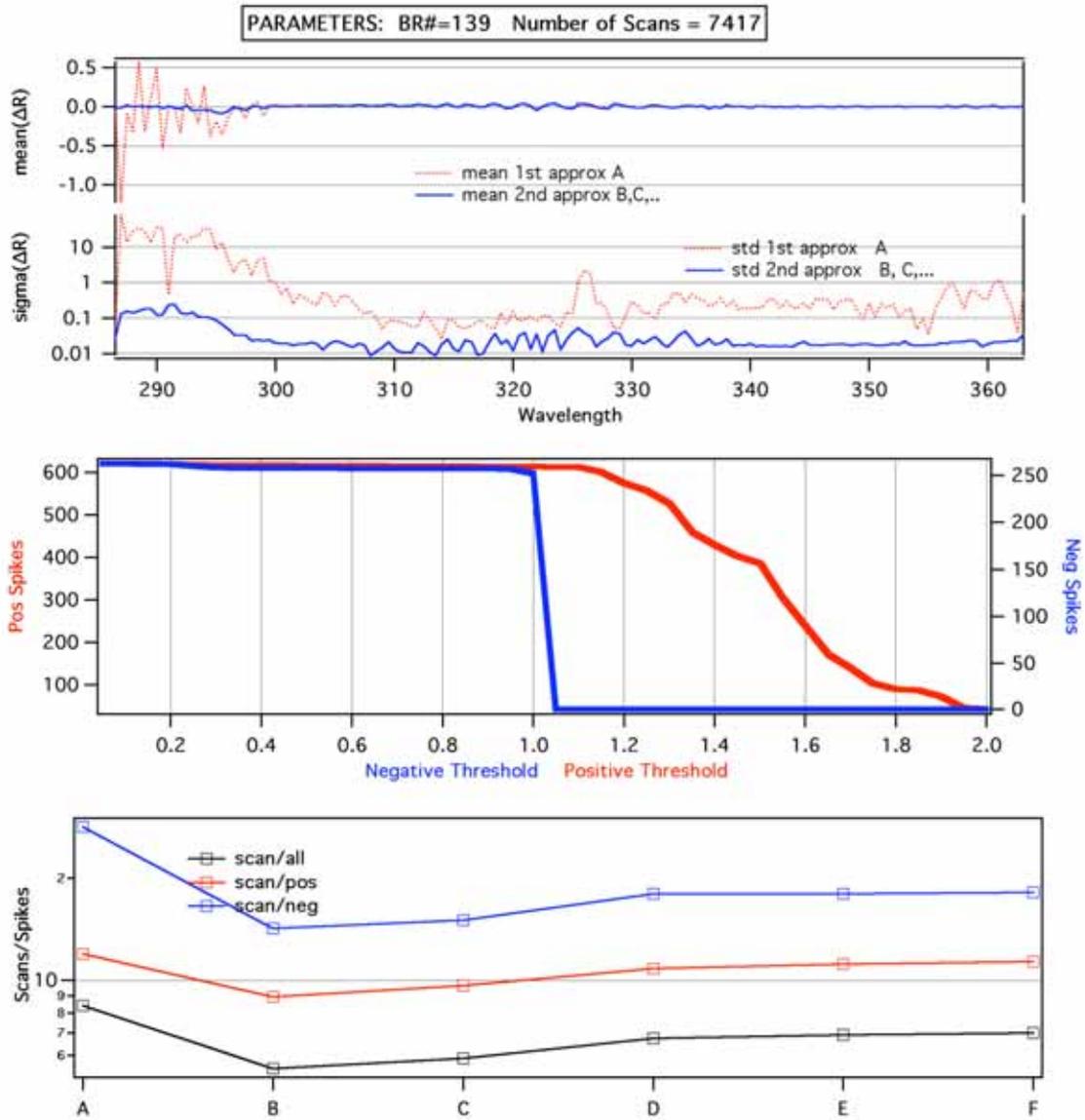
Graph 3 (3rd from top) scatter plot of spike irradiance versus spike ratio $M_{i+1} = r_i / r_i^c$ is presented. This graph and the previous one demonstrate that the choice of $R_{corrected} = 0.5$ was a good one.



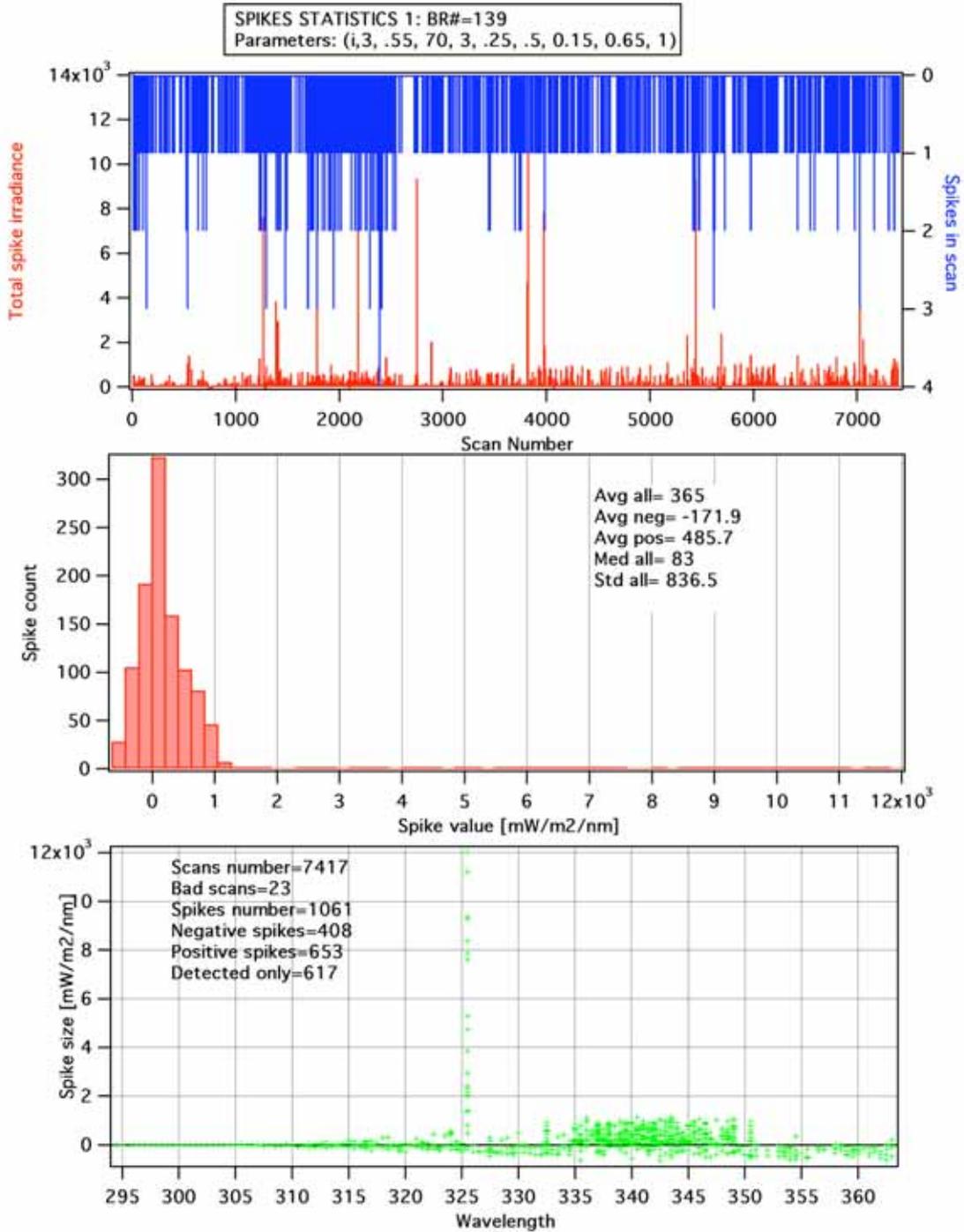
Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncld	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1214	1029	185	247.0	1451	0.07	9	3	345
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1481	1224	257	204.5	2634	0.06	9	3	802
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	1291	1112	179	234.6	1066	0.15	9	3	278
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	1015	985	30	290.8	1344	0.14	9	3	278
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	969	947	22	303.8	1389	0.14	9	3	278
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	931	919	12	315.3	1428	0.14	9	3	278

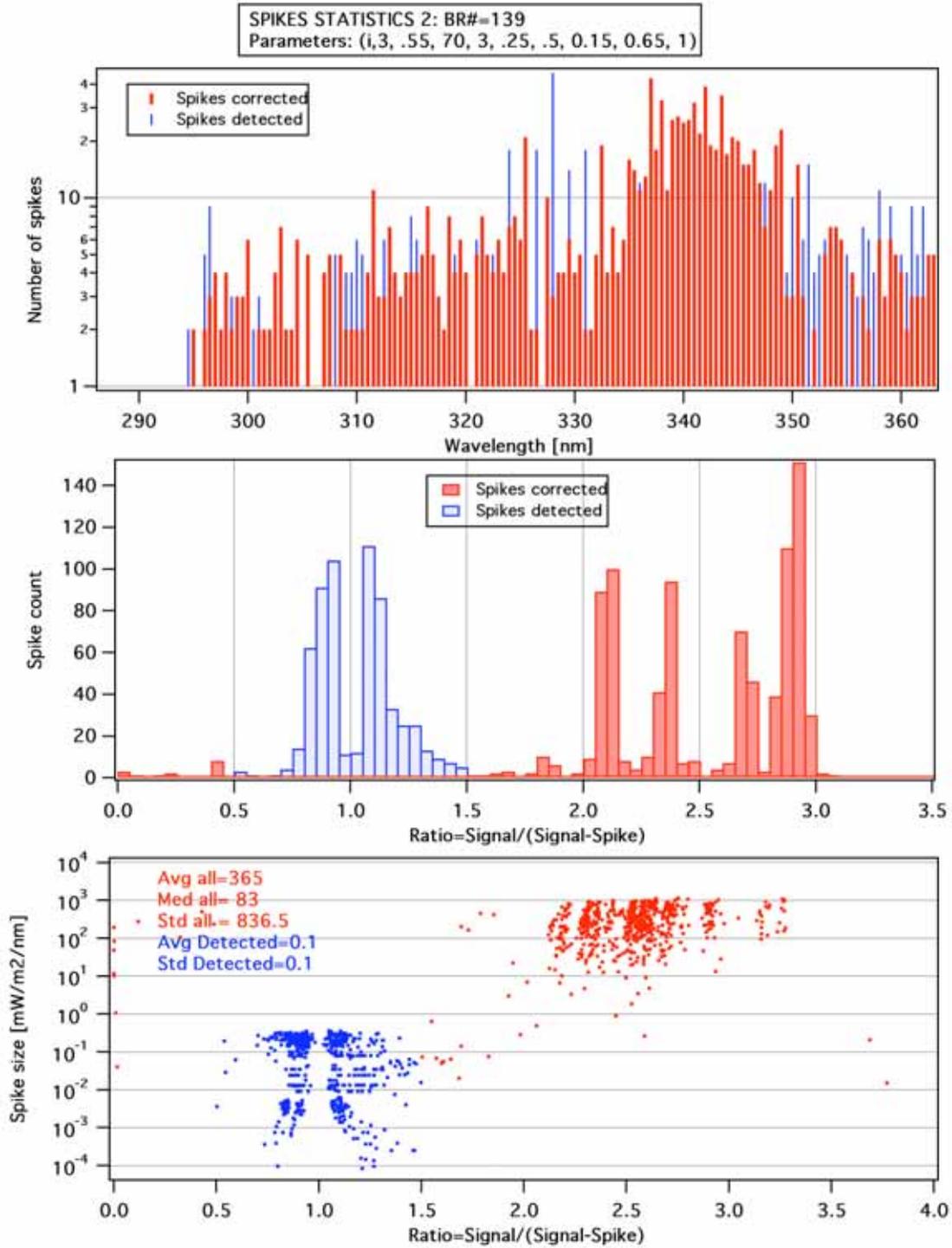


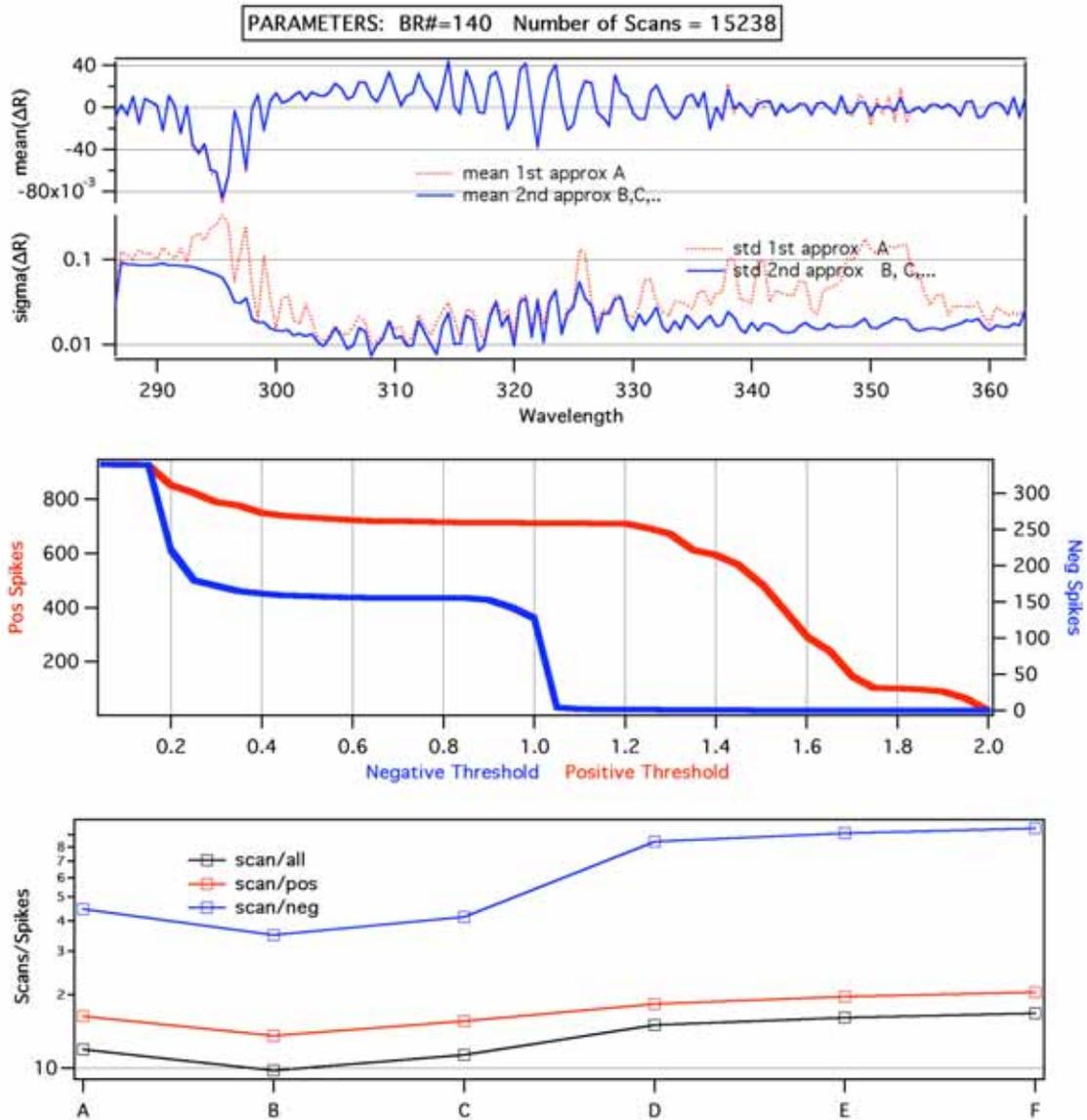




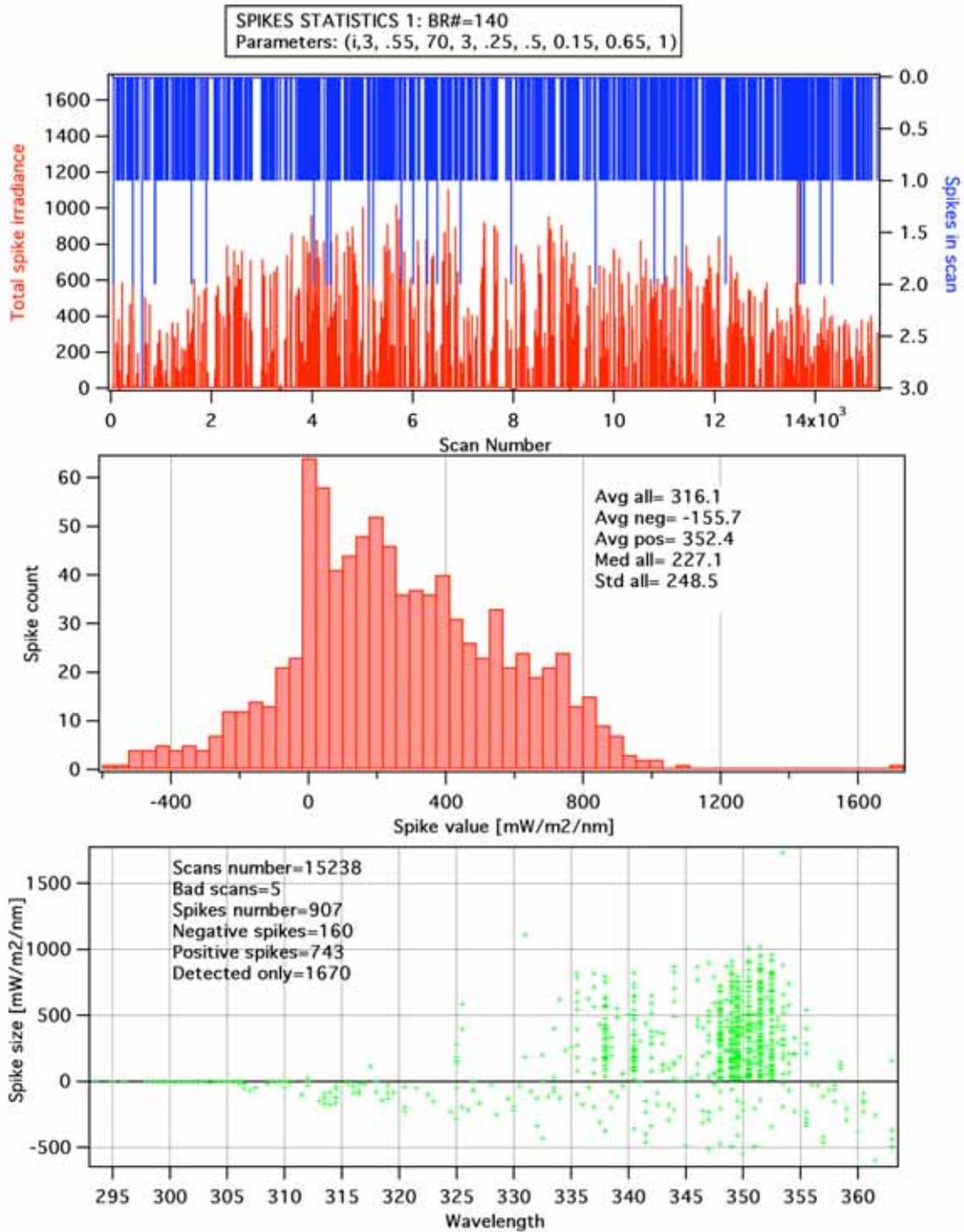
Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncld	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	882	620	262	411.8	0	0.00	107	0	9
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1353	831	522	290.1	1184	0.04	23	1	361
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	1262	769	493	311.0	415	0.12	23	1	111
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	1099	686	413	353.4	578	0.11	23	1	111
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	1076	664	412	360.5	601	0.11	23	1	111
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	1061	653	408	365.1	617	0.10	23	1	111

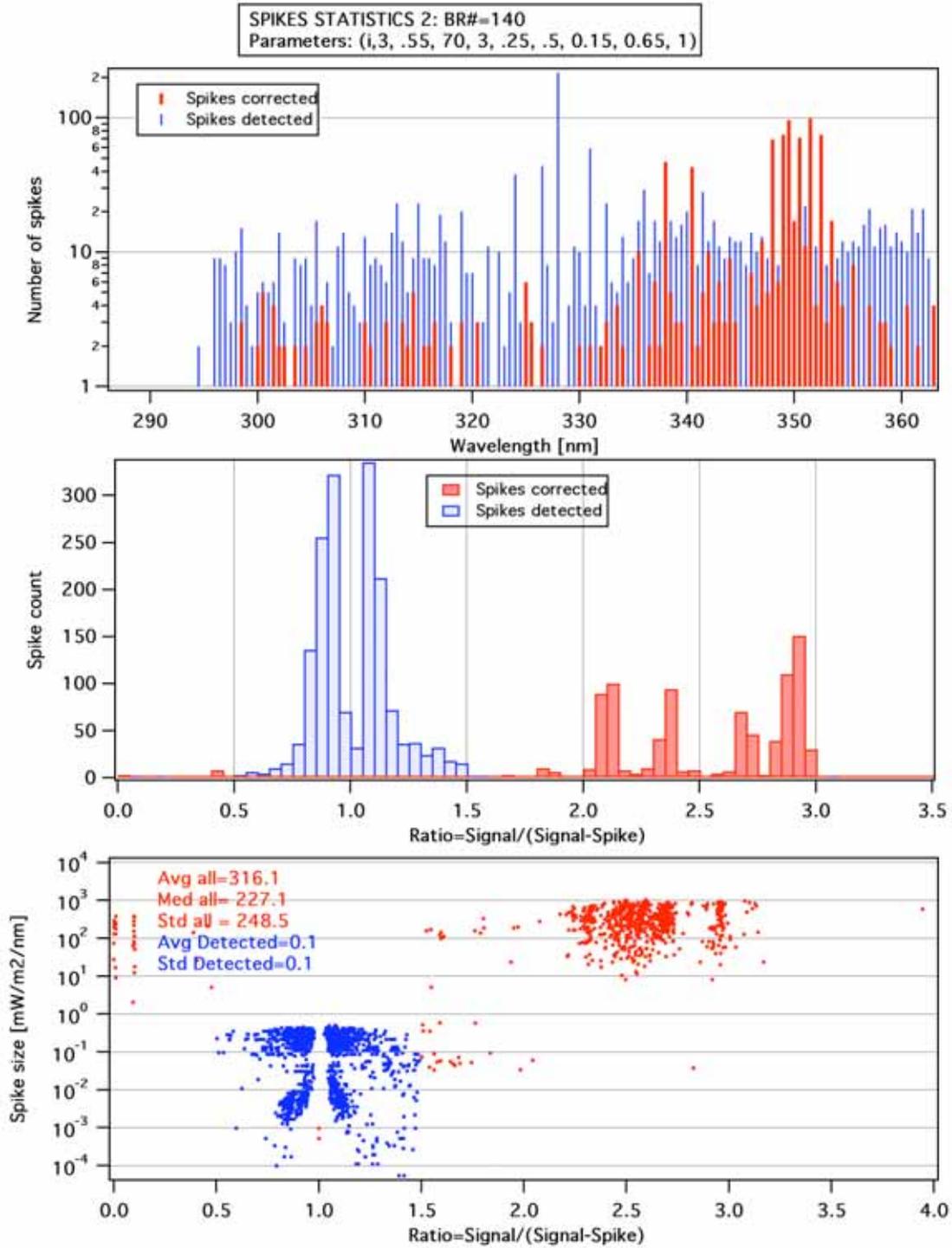


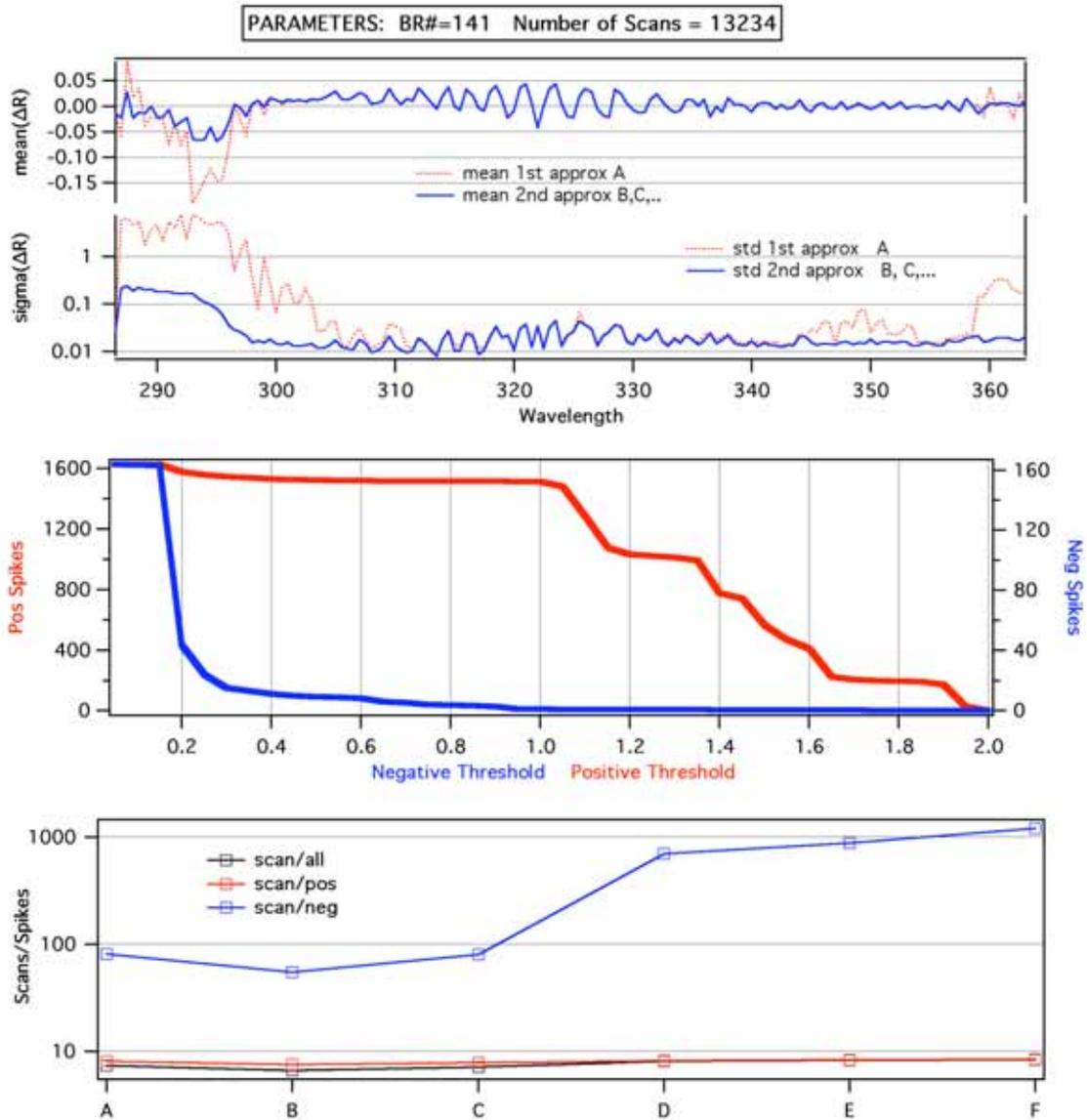




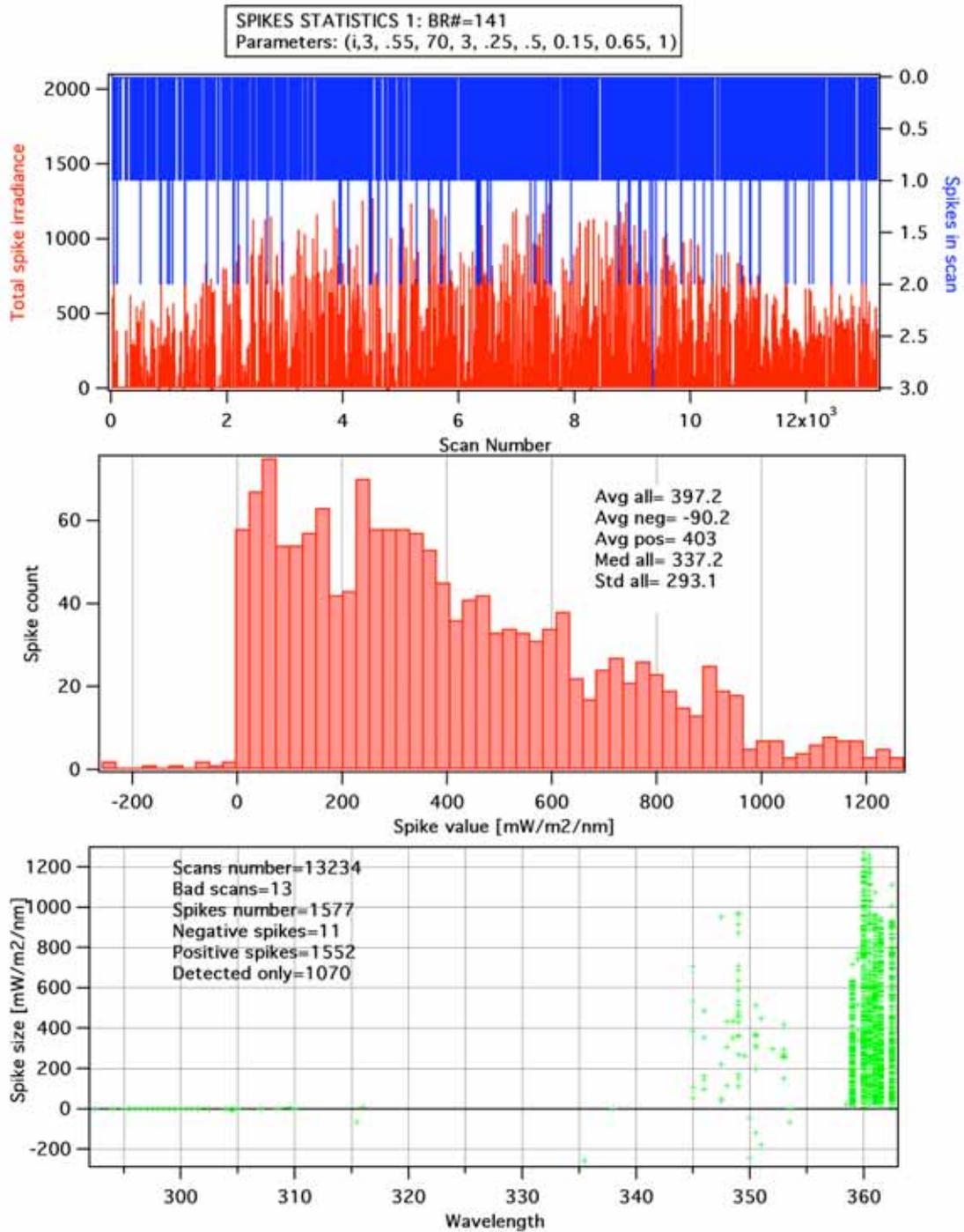
Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncld	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1277	936	341	232.7	572	0.11	5	2	299
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1558	1122	436	194.3	3084	0.06	5	3	1054
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	1346	978	368	224.9	1226	0.15	5	2	331
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	1015	834	181	286.8	1558	0.14	5	1	331
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	946	779	167	304.3	1628	0.14	5	1	331
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	907	747	160	316.2	1670	0.14	5	0	331

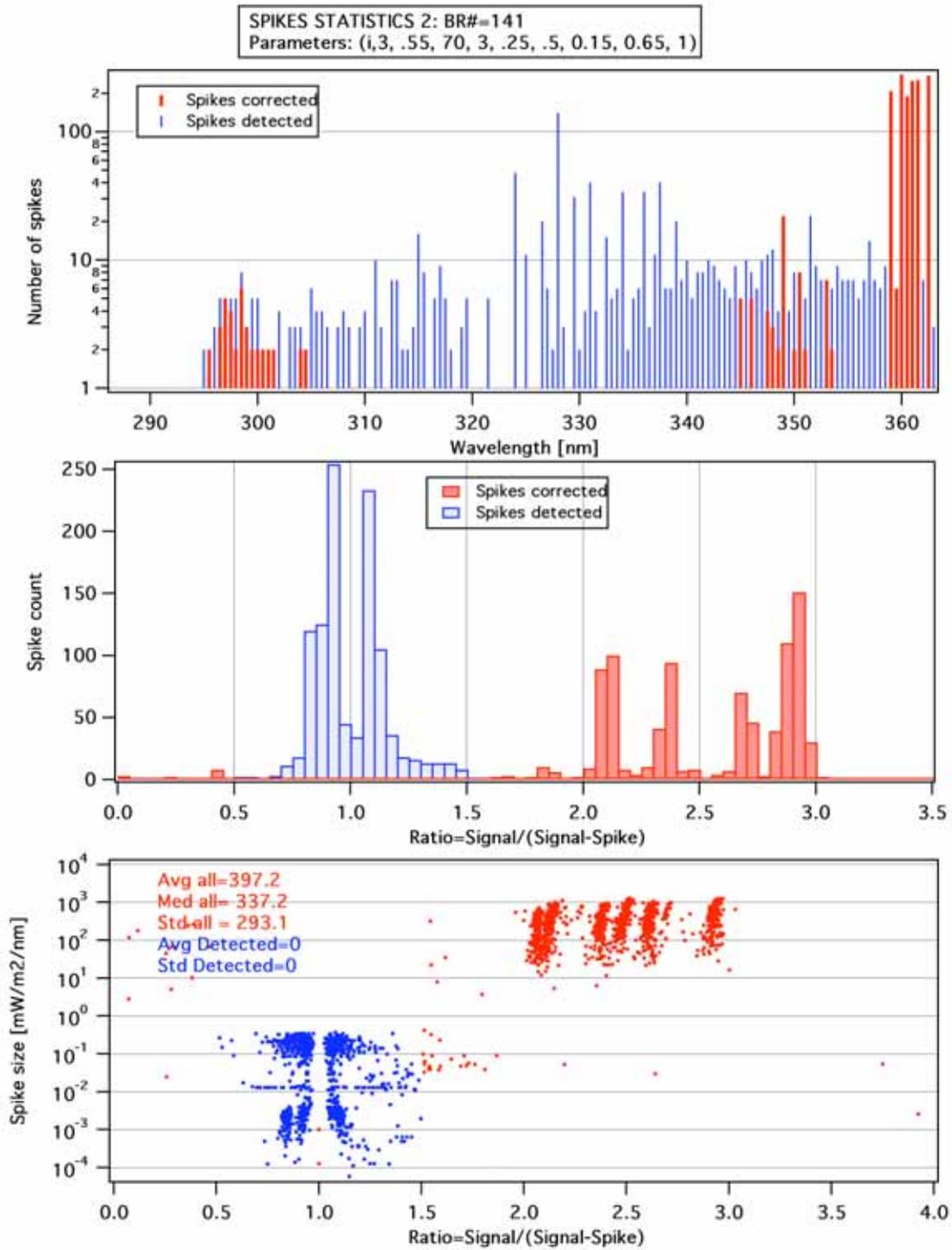


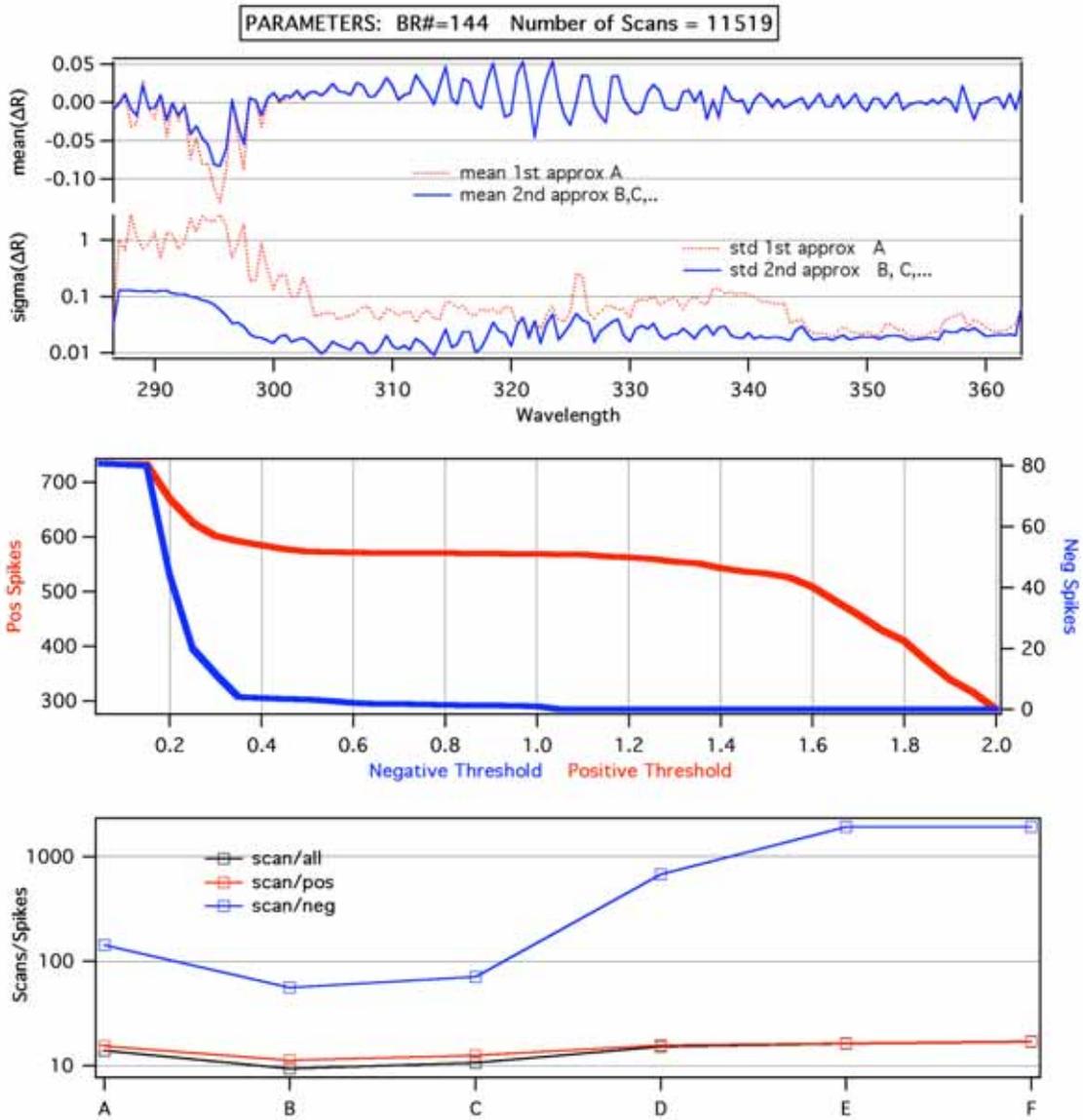




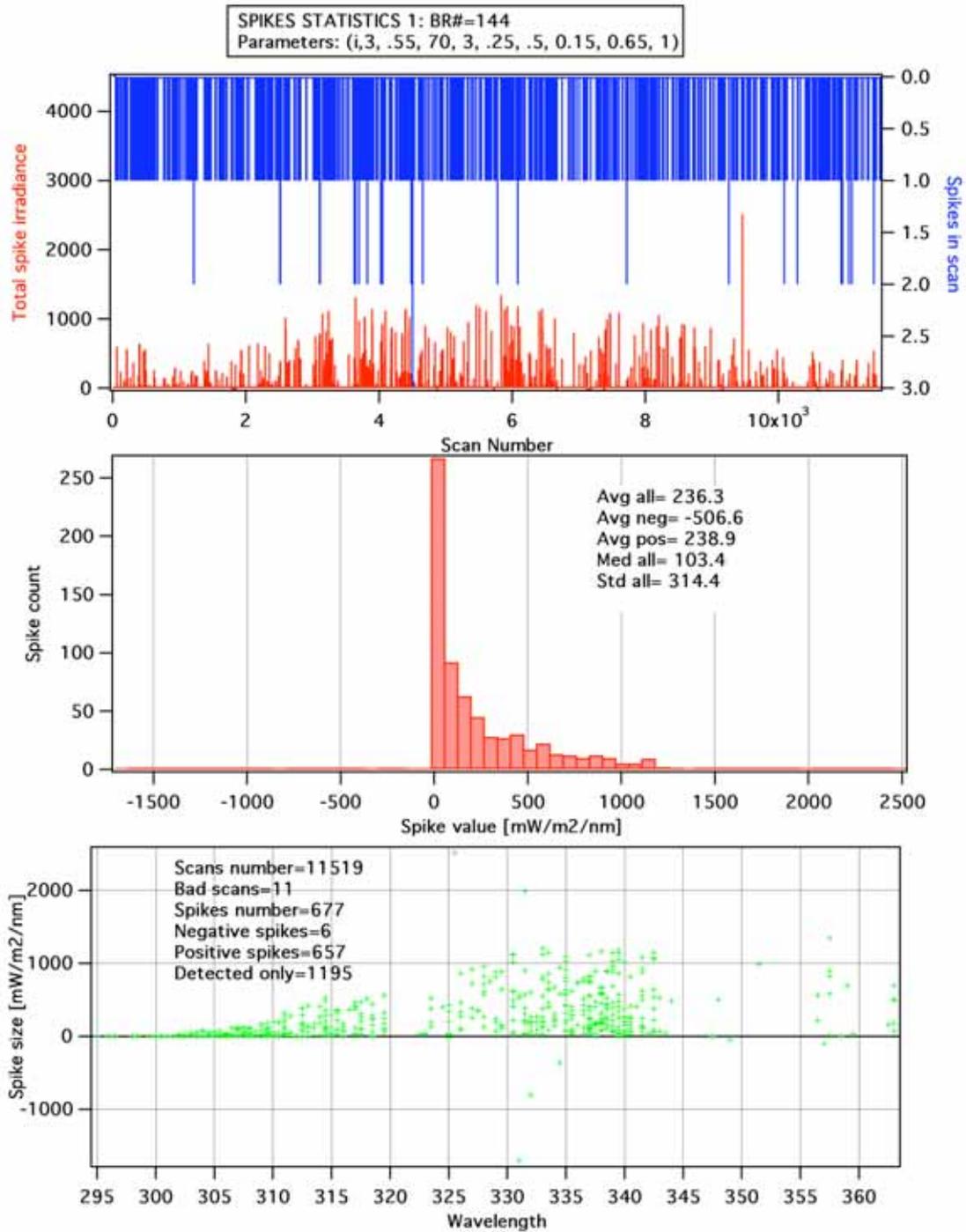
Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncld	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1798	1634	164	350.8	958	0.04	13	1	183
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	2000	1760	240	317.0	2109	0.03	13	1	694
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	1849	1683	166	342.9	796	0.09	13	1	144
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	1632	1613	19	385.2	1015	0.08	13	0	144
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	1602	1587	15	391.9	1045	0.08	13	0	144
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	1577	1566	11	397.3	1070	0.08	13	0	144

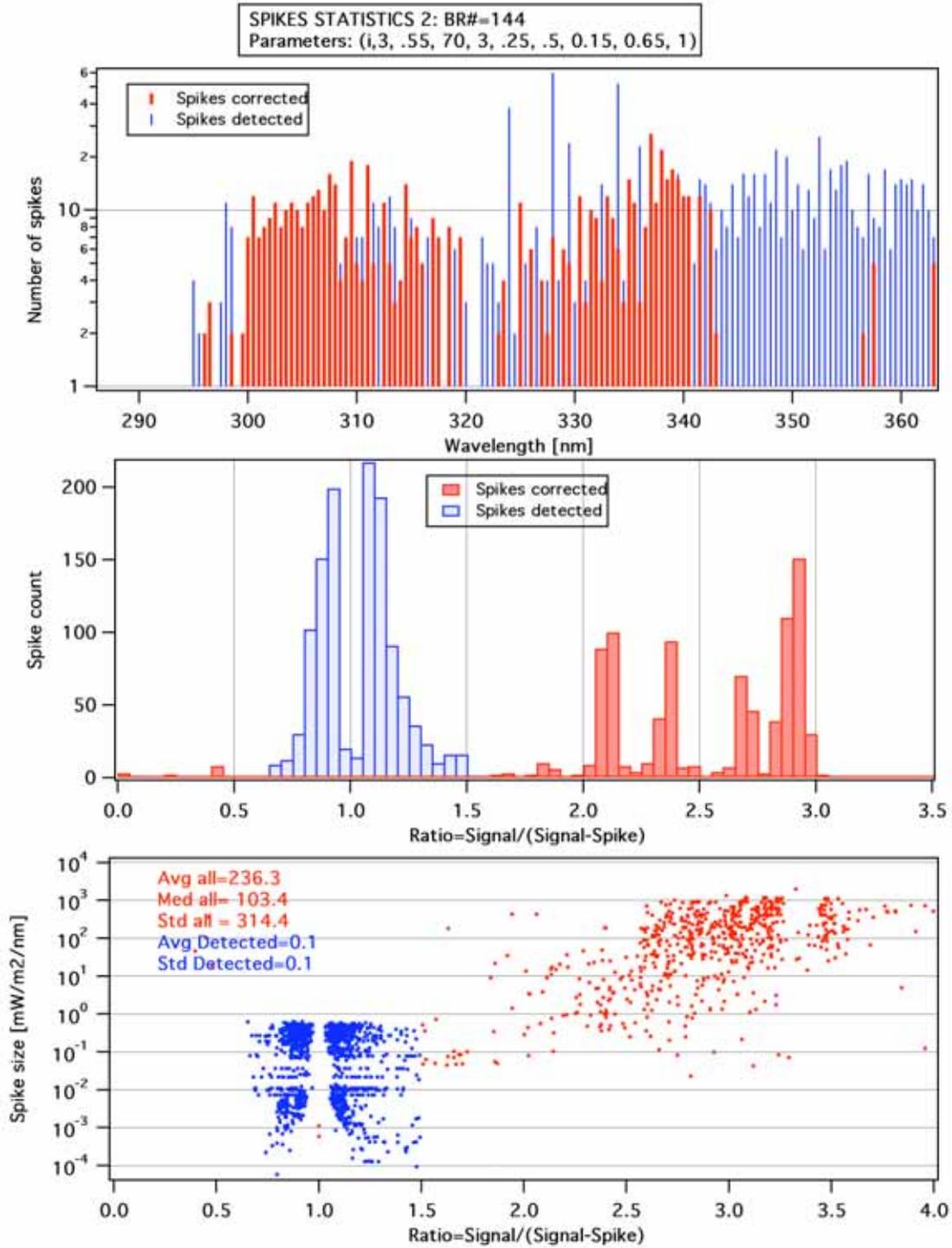


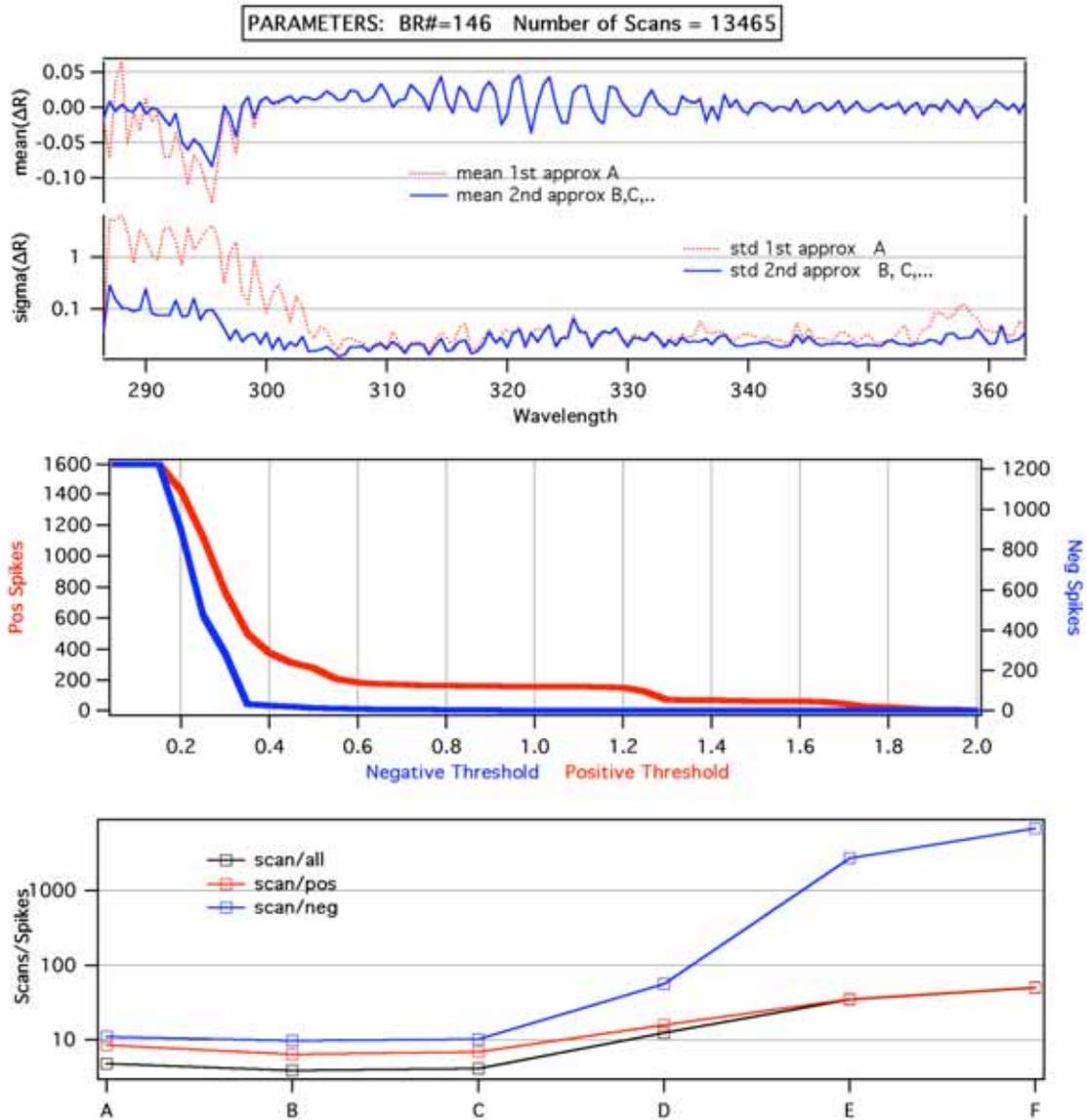




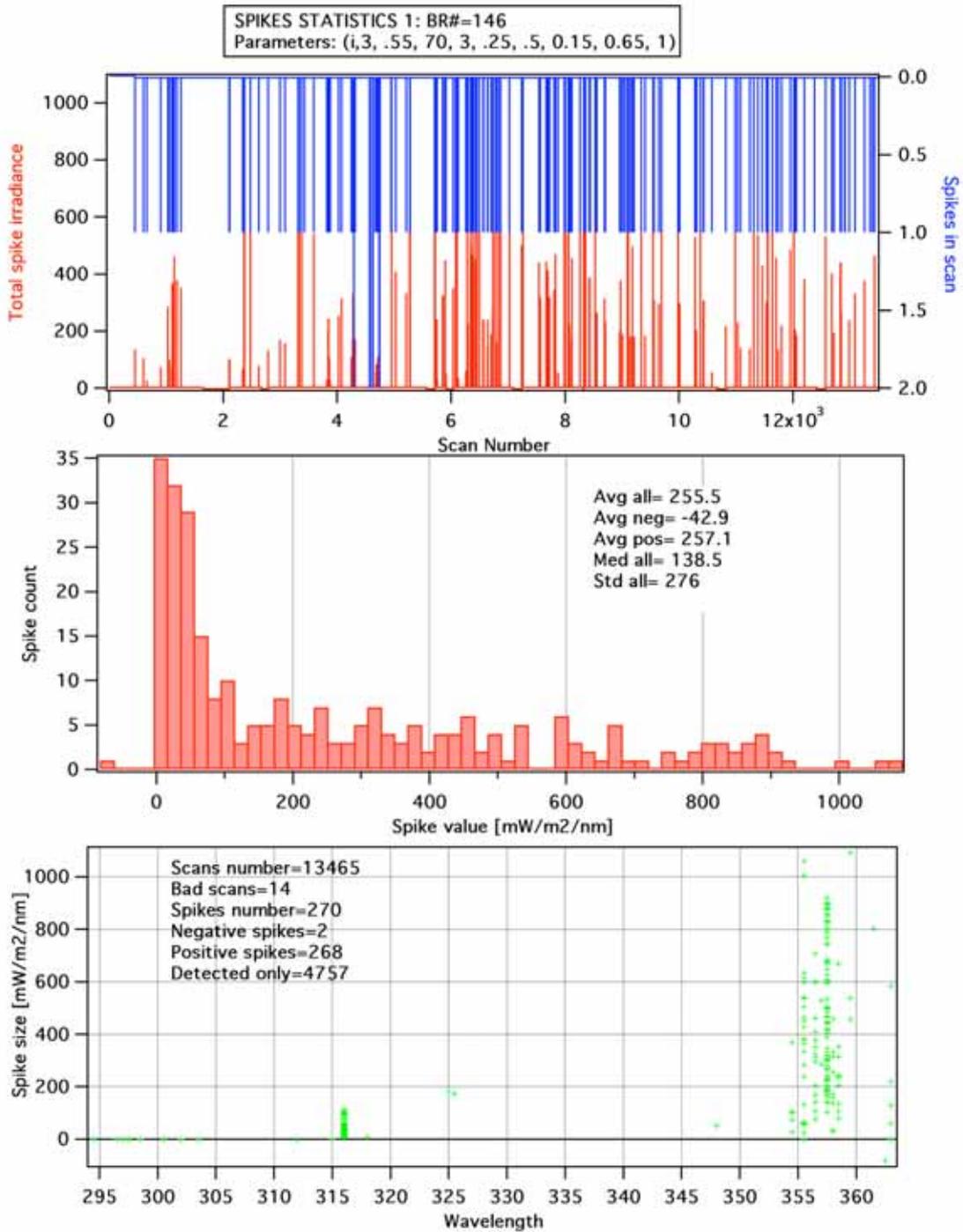
Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncl	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	827	746	81	203.3	566	0.07	11	0	172
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1228	1022	206	140.4	1999	0.08	11	0	767
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	1079	917	162	159.8	794	0.20	11	0	219
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	751	734	17	217.1	1121	0.18	11	0	219
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	708	702	6	227.2	1163	0.18	11	0	219
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	677	671	6	236.4	1195	0.18	11	0	219

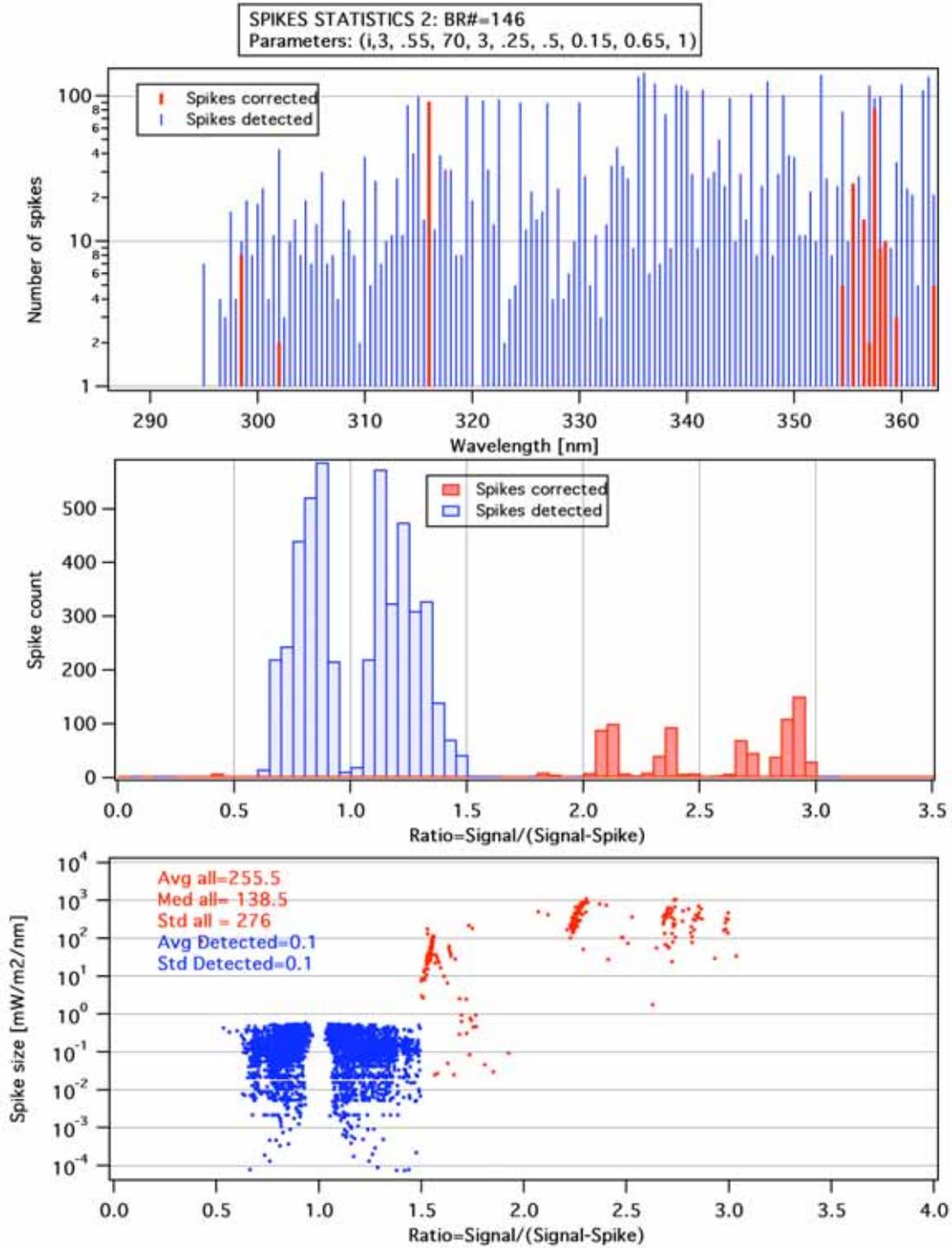


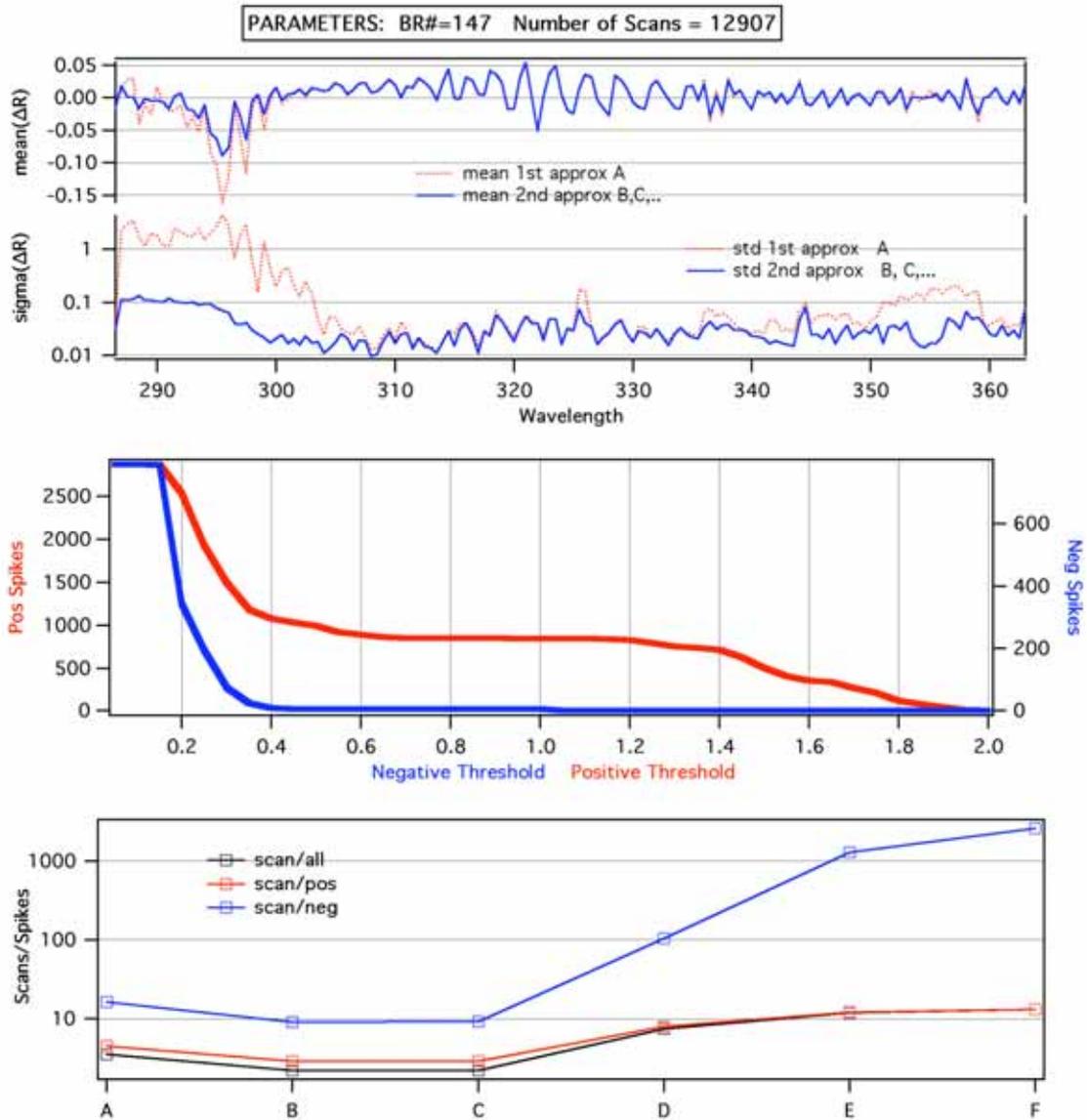




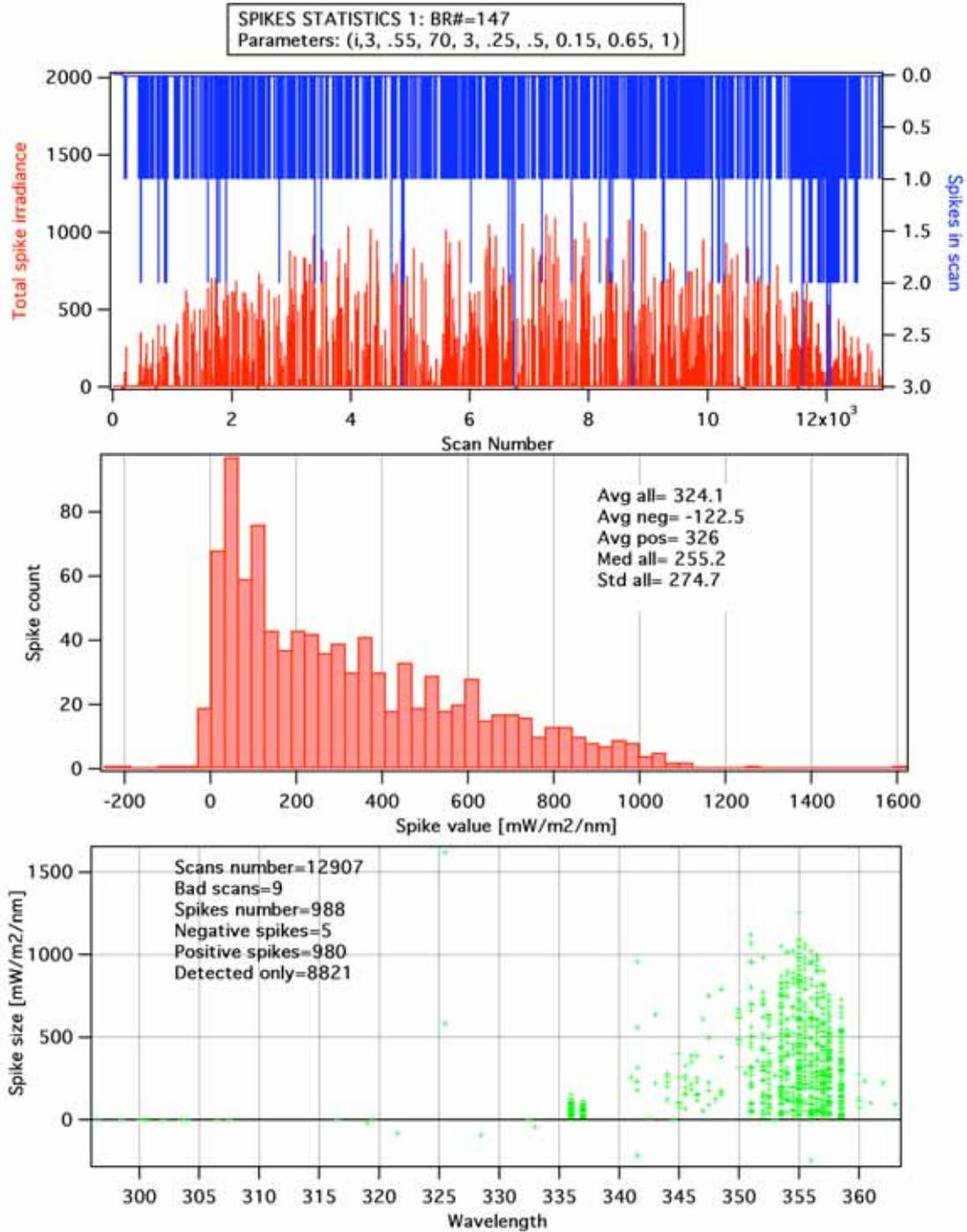
Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncld	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	2810	1587	1223	70.4	993	0.14	14	33	1085
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	3468	2086	1382	61.1	2168	0.14	14	66	1502
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	3270	1957	1313	64.8	1626	0.18	14	66	1228
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	1089	849	240	105.4	3939	0.16	14	0	1228
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	386	381	5	187.5	4642	0.15	14	0	1228
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	270	268	2	255.6	4757	0.15	14	0	1228

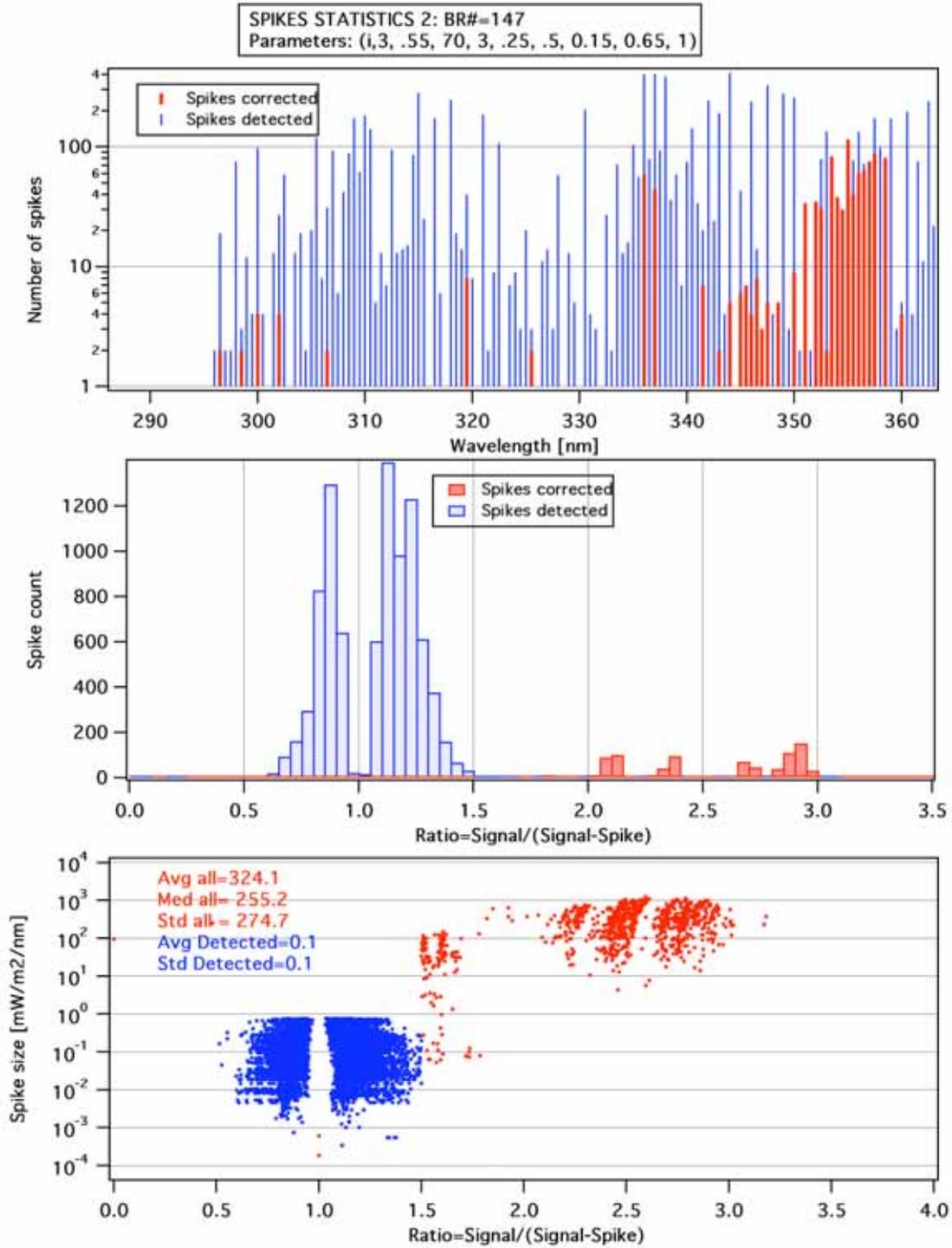


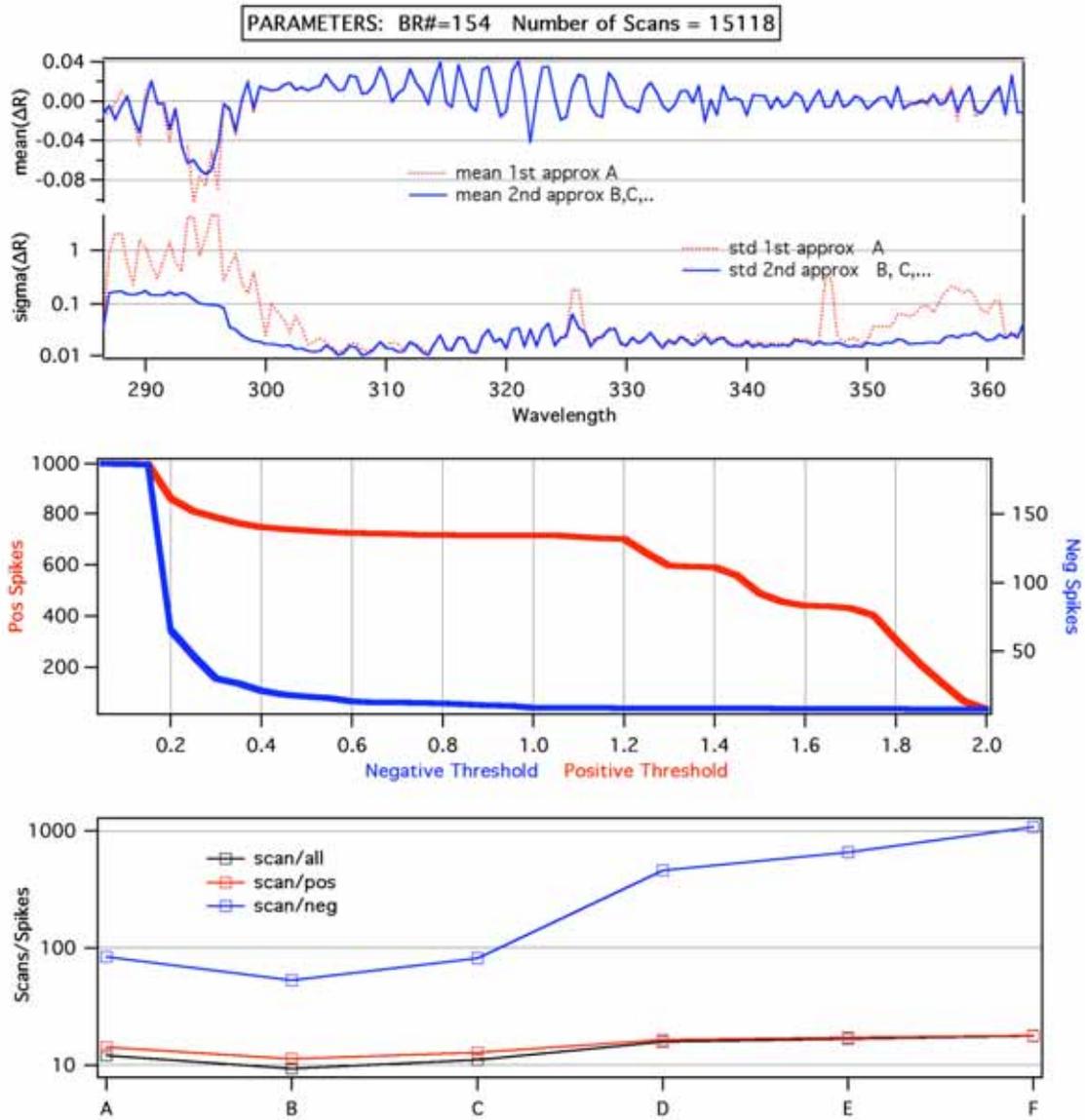




Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncld	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	3655	2866	789	108.9	1023	0.23	9	2	1818
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	5870	4449	1421	78.5	4057	0.20	9	2	3843
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	5838	4433	1405	78.9	3966	0.20	9	2	3703
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	1734	1610	124	201.1	8072	0.19	9	0	3703
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	1087	1077	10	298.2	8719	0.18	9	0	3703
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	988	983	5	324.1	8821	0.18	9	0	3703







Label	InputParameters	Nall	Npos	Nneg	Cavg	Ndtct	Davg	Nbad	Ncld	Ncoinc
A	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1248	1068	180	227.9	1629	0.24	4	0	392
B	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 0)	1628	1344	284	179.4	3048	0.22	4	0	970
C	(i,3, .55, 70, 3, .25, .15, 0.15, 0.65, 1)	1368	1183	185	213.4	1955	0.34	4	0	385
D	(i,3, .55, 70, 3, .25, .3, 0.15, 0.65, 1)	956	923	33	287.2	2367	0.34	4	0	385
E	(i,3, .55, 70, 3, .25, .4, 0.15, 0.65, 1)	902	879	23	302.1	2421	0.33	4	0	385
F	(i,3, .55, 70, 3, .25, .5, 0.15, 0.65, 1)	856	842	14	316.3	2466	0.33	4	0	385

