Discussion around Hyperion Data

David L B Jupp
April/May 2001

Introduction

Now that Hyperion data are here and being used, there are many things that need to be discussed and many experiences that can be profitably shared to prevent everyone falling into the same holes. This document is aimed to provide a discussion that invites other opinions and disagreement. It relies on information we at ANUTech have acquired from working with the data and from the privilege of being able to ask the experts at TRW questions over the last few months. As much as possible it uses and describes information accessible to you all on the oz_pi web site.

The first thing to keep in mind is that the data were released before the instrument team and the processing system had all come to an agreed optimum process and final product. This was done to get the science going as early as possible. There will therefore almost certainly be a number of versions of the data with the final version having best results in terms of calibration and minimal along track streaking. The data you get now WILL have streaks (pixels with gains relatively different in effect from their neighbours) and non-final calibration. So finding these effects is not a big discovery to make – rather an expectation.

TRW will soon be bringing out a mask letting us know which pixels in a line are likely to be off or unreliable. Later they will bring out improved products. The advent of the improved products will mean getting new data as the processes are nonlinear. It is important, therefore, that you record as much information about the data you get and processing you do as possible to ensure future traceability and re-processing and to allow direct comparisons to be made between results from different researchers.

None of you will get real Level 0 Hyperion data. One or maybe two will have a half way product which consists of data where dark current, smear and echo corrections have been made but where all the calibration gains are set to 1.0. The rest and (in the future all) of you will have the current Level 1 where a Cal file has been applied based on the pre-launch calibration and a solar calibration check on day 47. It was used to modify pre-launch calibration. This Cal file is called S047_Cal.dat and can be accessed among the files of the Frome data collection accessible through the oz_pi web area.

Given that you have data, the following discusses some of the issues of processing and choices of bands that you may wish to follow or simply think about. There are obviously many approaches being used around the EOC. My hope is that people in the ASVT will contribute their findings and details of techniques of processing to the discussion – as well as provide and show the final outputs. The total result will be better if it is a shared learning.

Wavelengths

For serious spectral research it is important to know is the effective wavelength of each channel/band. In common with all such instruments, the channels represent a wavelength band with a central wavelength and a bandwidth expressed as a central wavelength (cwl, λ_j) and Full Width Half Maximum (fwhm) of the spectral response function or distance between the points where the response function falls to half of its maximum value.

There is a nominal set in the header file where both cwl and fwhm of the bandwidth are recorded. The form of the bandpass function is taken to be Gaussian as defined by these parameters.

That is, if $L(\lambda)$ is the wavelength dependent radiance at the Hyperion entrance optics then the data for band j will be:

$$L_j = \int_{-\infty}^{\infty} \varphi_j(\lambda')L(\lambda')d\lambda'$$

where the bandpass function $\varphi_j(\lambda)$ is modelled as:

$$\varphi_j(\lambda) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(\lambda-\lambda_j)^2}{2\sigma_j^2}}$$

and

$$\sigma_j = \frac{FWHM_j}{2[2Ln2]^{1/2}}$$

Some discussion about this and its relationship with rectangular (eg *casi*) and triangular (eg Hymap) band passes can be found in the document FWHM_Notes.doc on the oz_pi area. For the results used to illustrate this document, a matrix was constructed mapping spectra on the 1 nm grid 350-2500 nm (2251 1nm bands) to Hyperion by approximating the Gaussian with the above formulae and re-normalising the 2251 columns to sum to 1.0.

You should check the values in the header with the file Hyperion_Cen_fwhm.dat on the oz_pi area. If it is not the same as the ones you have in the header you can replace the header values. These are band centres obtained by revisions of the published pre-launch wavelength calibration data and are more accurate. They should be used for plotting and reporting.

For Hyperion, the central wavelength and fwhm actually both vary across a line of pixels. The line is the formed by the dispersion of a slit across the detectors of the pushbroom which can lead to a non-uniform spectral response. That is, there is said to be a spectral “smile” in which both cwl and fwhm vary across the 256 pixels of a line. The data in the header are standardised by using the values for pixel 128.

There are two files on the web with the cwl and fwhm of the best fitting Gaussian to the measured Hyperion bands. The fit is very good and you will not have to go back to any other more basic files. Some of the plots are also on the oz_pi area for disbelievers.

The files are SpectralL0_RevA.dat and BandWidthL0.dat which are also available in the zip file Hyp_Band.zip. They are ascii files with 256 by 242 matrices of real numbers – one file for cwl and the other for fwhm.

The variations across a line are small but in hyperspectral work there are many atmospheric absorption bands that are narrow and deep and since the Hyperion fwhm is 10 nm on average (compare this with Hymap which is 15nm) there will be obvious effects that vary with pixel related to the “smile”. For some atmospheric effects it has been noticed that a shift of 1/20 of the fwhm creates a significant effect. For Hyperion this translates to 0.5 nm, which makes it necessary to take the “smile” into account over the full range – but especially in the visible where a range of 3.5 nm occurs at the blue end.[1]

Taking channels 13, 22, 29, 46, 144, 196 as example bands (they can be thought of as surrogates for the Landsat ATM bands and can therefore be used to obtain more “familiar” plots from Hyperion), we can see how smile changes with both wavelength and pixel position.

The statistics across a line (average, minimum, maximum and range in nm) for these bands are:

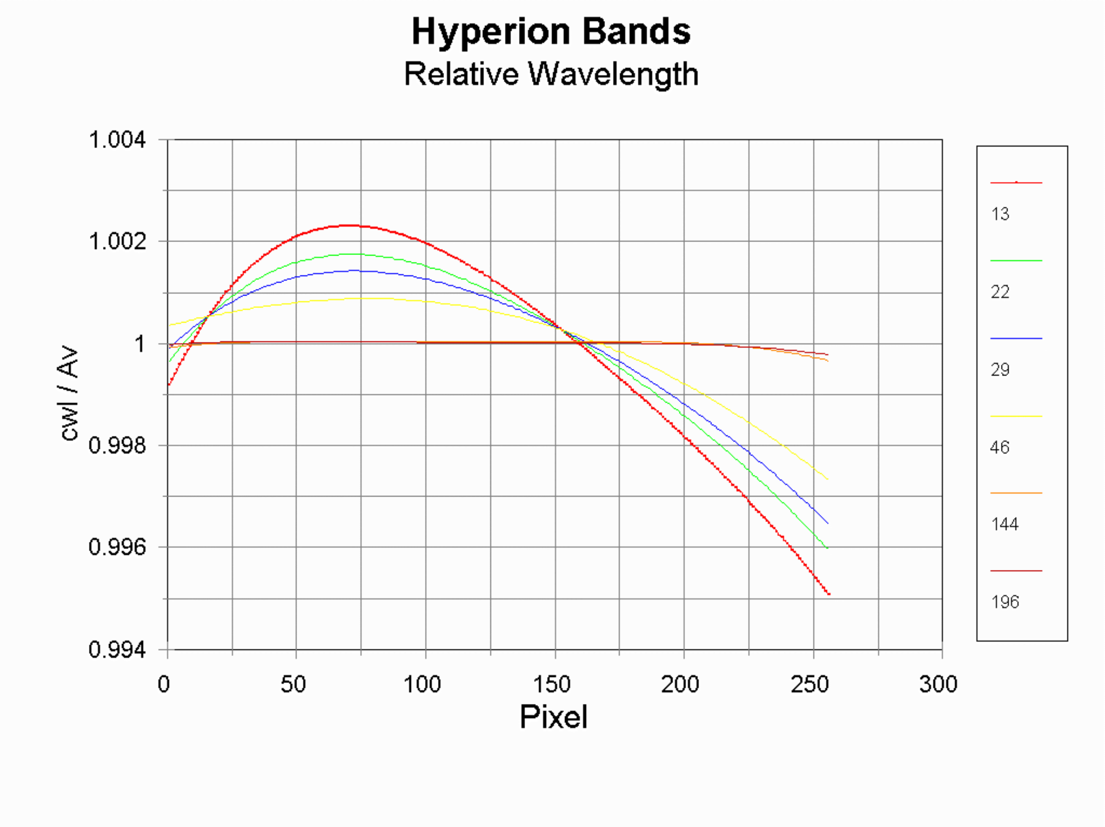
“Landsat” Hyperion Band	**Avge**	**Min**	**Max**	**Range (nm)**
13	476.27	473.93	477.37	3.44
22	568.05	565.76	569.05	3.29
29	639.45	637.19	640.37	3.18
46	812.84	810.66	813.55	2.89
144	1588.42	1587.88	1588.49	0.61
196	2113.04	2112.58	2113.14	0.56

Which shows how the smile is most active in the VNIR and almost non-existent in the SWIR. Obviously in the VNIR a range of 3 nm must rate as highly significant for the application of atmospheric correction or modelling. It implies (for example) that the matrix mentioned above depends on pixel number.

Actually, as illustrated below in the graph of the relative cwl (cwl over average cwl) for these bands across a line, the “smile” is really a “frown”. But, fortunately, pixel 128 is about central in the variation.

---

[1] It has been claimed that the underlying resolution for atmospheric models and transmission codes in this case will be 0.1 nm. This is close to the limit of Modtran 3 in the visible range.



Nevertheless, initially it is best to accept the nominal cwl values given in the header files that come with the data for your work and plotting[2]. The main effects of the frown will be to make some of the atmospheric absorption areas unstable so below I will suggest zones to exclude from processing (especially for MNF, PCA or Classification – or ratios) unless you know what you are up to (normally for atmospheric correction).

Useable Channels & Feature Detection

The Hyperion data you will receive consist of 242 channels with 256 pixels in a line and a certain number of lines of data. Of these 242, 200 channels have been selected in the Level 1 data product as the primary useable channels. The reduction from 242 to 200 comes about partly because a number of the detectors were "not illuminated" in the arrays and partly from the fact that in the 242 detectors from the two arrays (70 in the VNIR array and 172 in the SWIR array) there is a region of overlap. The sensitivity of the materials falls well off optimum in some places of the overlapping area so some of these have been eliminated.

This is done now in the Cal file so in the files you have got and will get only the following bands should be active (others set to zero):

---

[2] The option of resampling to a standard set of wavelengths at the same time as correcting for readout on the geometry is discussed briefly later.

VNIR Range: Bands 9-57 (436nm-926nm)
SWIR Range: Bands 75-225 (892nm-2406nm)

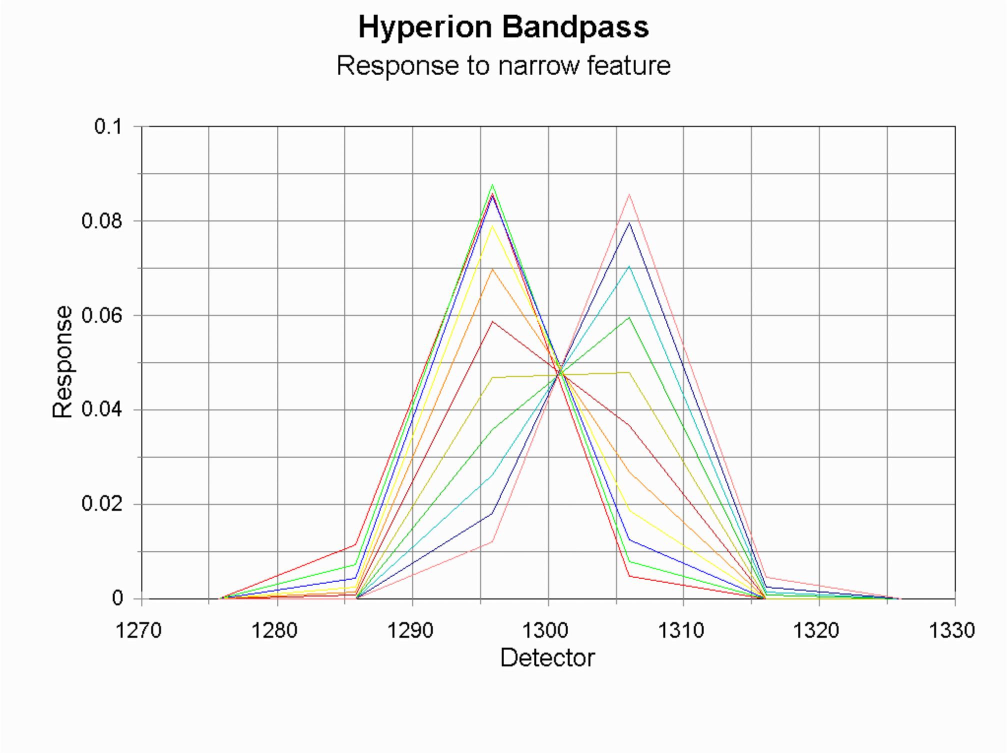
This adds up to 200 essentially useable channels with 196 unique wavebands with 4 channels [54-57, 75-78] remaining in the overlap.

It is not planned to distribute a 200 channel file or 196 channel file. However, if you ever get one either from NASA or an SVT member, the above are the selections from the original bands. The original bands and band numbers are used in the remainder of this document.

Even within these ranges there will be some non-illuminated detectors scattered occasionally through the lines and differently from line to line. So, beware. There are not too many, but there are some and some classifiers and PCA/MNFs tend to find the effects of such niggly things. Later, a pixel mask will be provided which identifies these and data may have them filled by neighbour averaging or you can fill them in from the mask using neighbourhood pixel values.

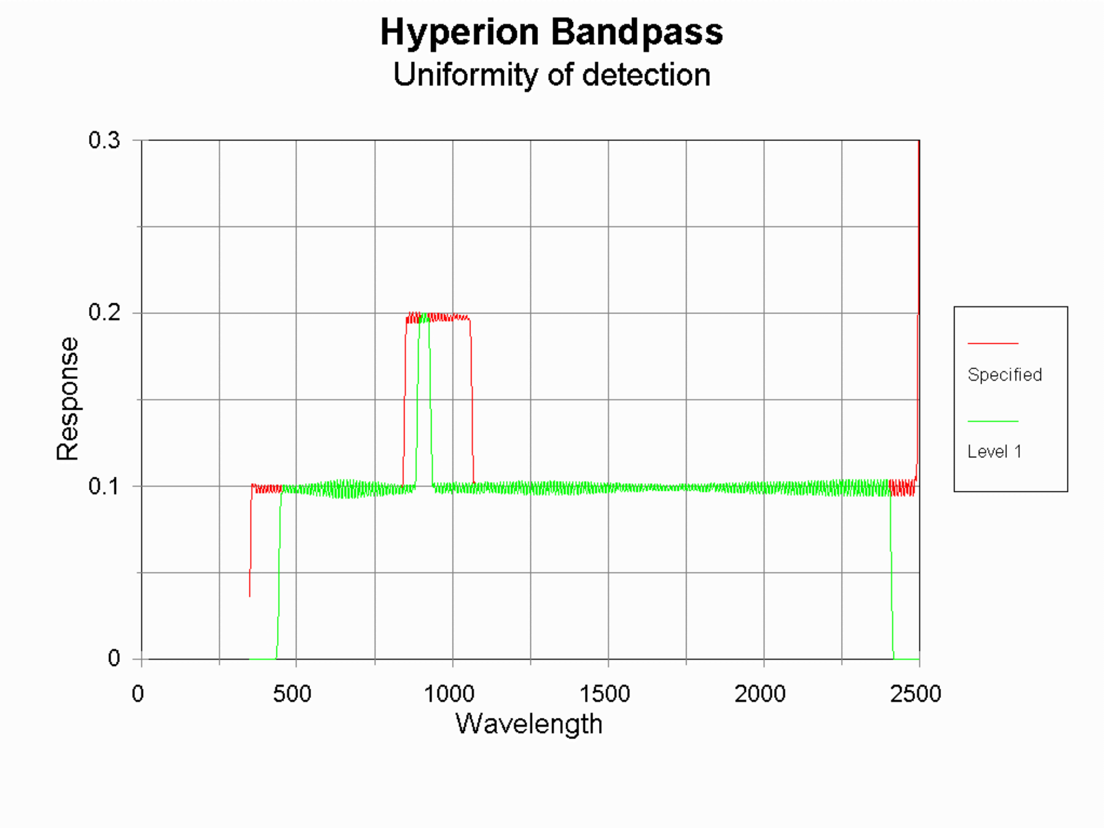
It is useful and informative to investigate how well the original design and final selection of bands cover the spectral regions of interest and resolve spectral features of interest. In the original design of Gaussian bandpasses separated approximately at a step of one fwnm it can be shown that the presence and magnitude of a spectral feature will always be equally represented in the data as it moves spectrally – although it may be distributed differently across the wavebands.

The actual Hyperion data are not exactly this design but are close and can be used to illustrate the point. Suppose a spectral feature is concentrated into a 1 nm width and varies in position from 1295-1305 nm in steps of 1 nm. The Hyperion band response would be:



This shows how shifts of concentrated features of even 1 nm could be detected if the SNR is high enough and how the "strength" of the signal remains the same albeit distributed across a number of bands. This means that the Hyperion design provides a very discerning tool for spectral studies.

Another way of illustrating the uniformity of response is to sum all the responses at each wavelength. Instead of doing this mathematically, we can simply sum the values of the approximate response set we have constructed on a 1 nm spacing of wavelengths. The sum plots out as:



The plot shows the effect of reducing the bands and clearly the capacity to track spectral shifts and represent features without loss is high. The jump to double the value is the overlap area. There is also a variance due to the discrete approximation and "smile" on the 1 nm grid.

Note, however, that tracking features at this level requires close attention to the spectral "smile" as well as high SNR.

Radiance Units and Atmospheric Absorption Effects

The data you get should be in units of (SI units) W m^{-2} μm^{-1} sr^{-1} times 10^3. However, since changing to nm, cm and mW all just shift by powers of 10 you can to some

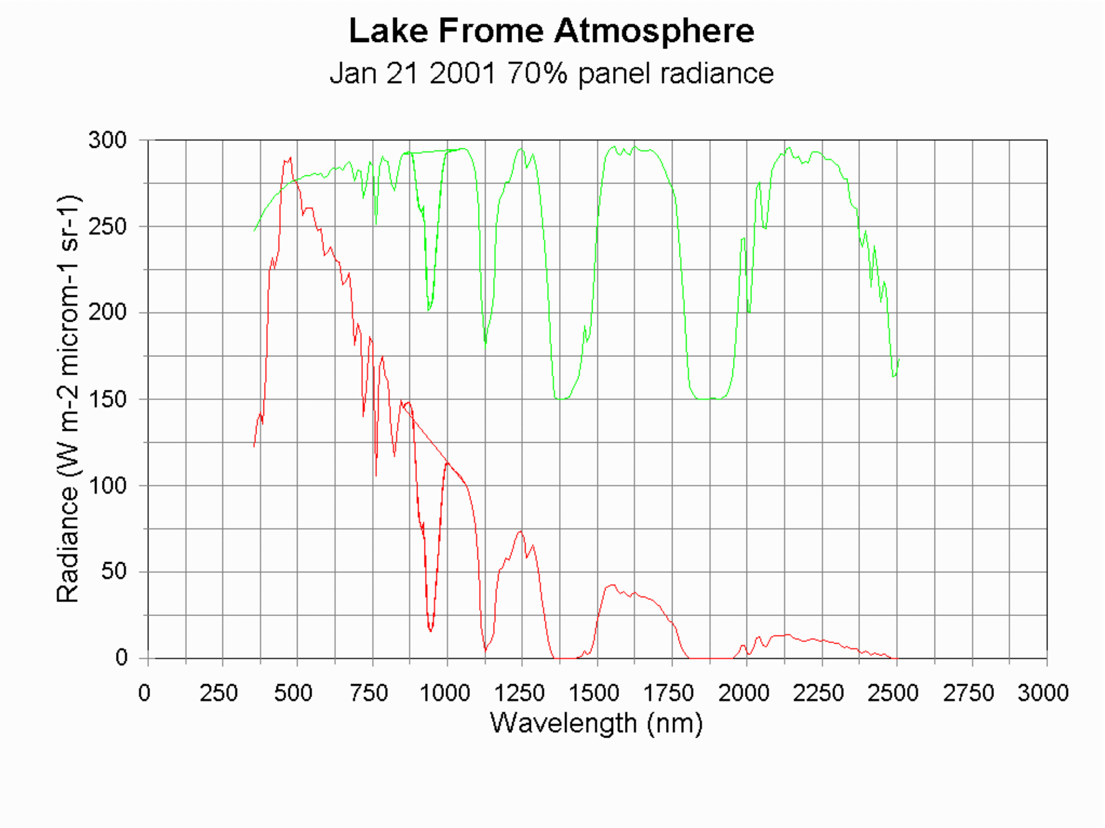
---

[3] There are variations here and also some discussion going on about a possible times 10 in the VNIR and times 100 in the SWIR. Watch for news! Many of you will have times 100 data. The best plan is to model the expected radiance as a check.

degree decide for yourself the choice of these units and the multiplier. SI units, however, should be used for publications.

There were a few specific early data sets (Frome, Fitton and maybe Tumbarumba) which are 10 times the SI units. Then products have generally been 100 times the SI units and finally the current Level 1 products are coming out as 10 times SI units. There is some talk of having a different multiplier in the VNIR and SWIR. So, if you have any trouble try both or call us and check the source and date of the supplied data.

The following graph shows the expected radiance at sensor for a target of flat, gray and Lambertian reflectance of 0.7 or (70%) on January 21, 2001 at Lake Frome. No detectors have been excluded here. The scale on this plot may help you see if the units you have used are OK. Also plotted (scaled) is the atmospheric transmission in green (150 is zero and 300 is 1).



Since the flat target spectrum modelled here has no "colour" there are quite a few obvious features about the plot which will be common to all of your data and are not related at all to the spectral type of the target on the ground.

The main structure is an imprint of the distribution of the solar radiation (like a black body of temperature about 6000K with Fraunhoffer lines) modified by atmospheric absorption and scattering.

There are three major water vapour absorption areas (with "fwhm" at about 1350-1480 nm, 1800-1970 nm and > 2480 nm) and some lesser but still significant ones between 930-960 nm and 1115-1150 nm plus features due to O_2 (a major band at 760 nm and a smaller band at 690 nm and a really little one at 1260 nm) and CO_2 (1950-

2050 nm and 1600-1610 nm) and various gases with minor effects in this range. Two smaller water absorption features flank the O_2 line at 760 nm as well. Most of you probably know all this. However, just in case…

The issue with these areas is that there is very little signal from the ground surface and also they tend to be unstable with regard to changes in cwl. It is for this reason that Hymap avoids them. However, Hyperion basically takes no special account of the location of the atmospheric "windows" but spaces the bands about 10 nm apart between 300 nm and 2500 nm without fear or favour.

I would therefore suggest that people trying out classifications or MNFs or other transformations with a view to seeing what Hyperion says about the surface could also do well to disregard the following channels (in addition to those listed above):

Wavelengths (nm)	Channels	Comment
760	41	Oxygen "notch"
930-960	79-81	Water Vapour
1115-1150	98-101	Water Vapour
1350-1480	121-133	Water Vapour
1800-1970	165-182	Water Vapour
> 2480	(already zapped)	Water Vapour

Some of you may prefer to be even more bold and eliminate more of the channels on the roll-off from the major absorption bands. For example, the band 930-960 could be broadened to 890-1000 nm to remove the water vapour effect. However, the best choice depends on what you want from the data and where the important features are in the ground targets. Good luck.

With the current calibration you will find issues at about 900 nm (channels 54-57, 75-78) due to the cross-over between the detectors. At this stage they may not match after calibration. This actually creates some problems for the atmospheric models that use the 930-960 nm Water band region to estimate water vapour since one of the reference bands falls into this region.

Maybe since the above is negative (what to leave out) I will list some channels to USE as follows:

Wavelengths (nm)	Channels	Name
446-750	10-40	VNIR_1
770-926	42-57	VNIR_2
902-920	76-78	Overlap SWIR
970-1105	82-97	SWIR_1
1160-1340	102-120	SWIR_2
1490-1790	134-164	SWIR_3
1980-2406	183-225	SWIR_4

Personally, I would prefer to drop the Overlap SWIR leaving 156 channels to use for exploratory data analysis for earth surface features and discoveries.

Note that the central wavebands of ALI correspond approximately to the Hyperion channels and the major groupings listed above in atmospheric windows:

Hyperion Band	ALI cwl	ALI Band	Range
10	441.6	1p	VNIR_1
14	484.8	1	VNIR_1
22	567.2	2	VNIR_1
31	660.0	3	VNIR_1
44	790.0	4	VNIR_2
51	865.6	4p	VNIR_2
110	1244.4	5p	SWIR_2
149	1640.1	5	SWIR_3
207	2225.7	6	SWIR_4

Nevertheless, SWIR_1 seems actually to be a very interesting area. Maybe the next Landsat can have 10 channels if we find out really how useful SWIR_1 is by using Hyperion?

Atmospheric Correction

Having selected channels and looked at data, plotted them and transformed them, you must decide if the image needs to be atmospherically corrected. If you are doing classification of a single image and/or looking at separability with CVA or exploring with MNF it really is not necessary to atmospherically correct. Actually, it is not necessary even for the data to be calibrated – and there are some valid reasons why the results can be better if the data are neither calibrated nor atmospherically corrected – especially the latter. However, if you want to compare the images with site spectra and library spectra or apply models or compare images at different times then maybe it is quite necessary.

This is not the place to go into the many options. That is a bigger discussion and one that we should start at ASVT and SVT meetings. But I will mention one option some of you can (and already do under various names such as “empirical line”) employ to very good effect.

If you have a number (two minimum, three better and more even better) of site spectra or PIFs that you can or feel you can identify in an image, you can apply a general rough correction as described below. In fact, this analysis is a useful thing to do even if you are going for full model-based atmospheric correction using ancillary data such as irradiances and radiosondes.

The method assumes you have a number (M) of known or assumed known spectra (i=1,M) at N wavebands (j=1,N). Since Hyperion has a very narrow angular width and constant pointing it is reasonable to assume the atmospheric components are slowly varying. Reasonable, but not quite right, of course, due to spatially varying changes in the atmosphere and the spectral “smile”.

Anyway, writing this out, it would mean to a reasonable approximation and assuming the background effect is slowly varying that the radiance at sensor (L_{ij}):

$$L_{ij} = a_j \rho_{ij} + b_j$$

Taking means over the examples:

$$\bar{L}_j = a_j \bar{\rho}_j + b_j$$

So, we have simple solutions for the a_j and b_j as:

$$a_j = \frac{Cov(L_j, \rho_j)}{Var(\rho_j)}$$

$$b_j = \bar{L}_j - a_j \bar{\rho}_j$$

Basically, *b* should approximate the path radiance and *a* should be like irradiance times transmittance over pi. Obviously a simple linear correction for each channel follows as:

$$\tilde{\rho}_j = \frac{L_j - b_j}{a_j}$$

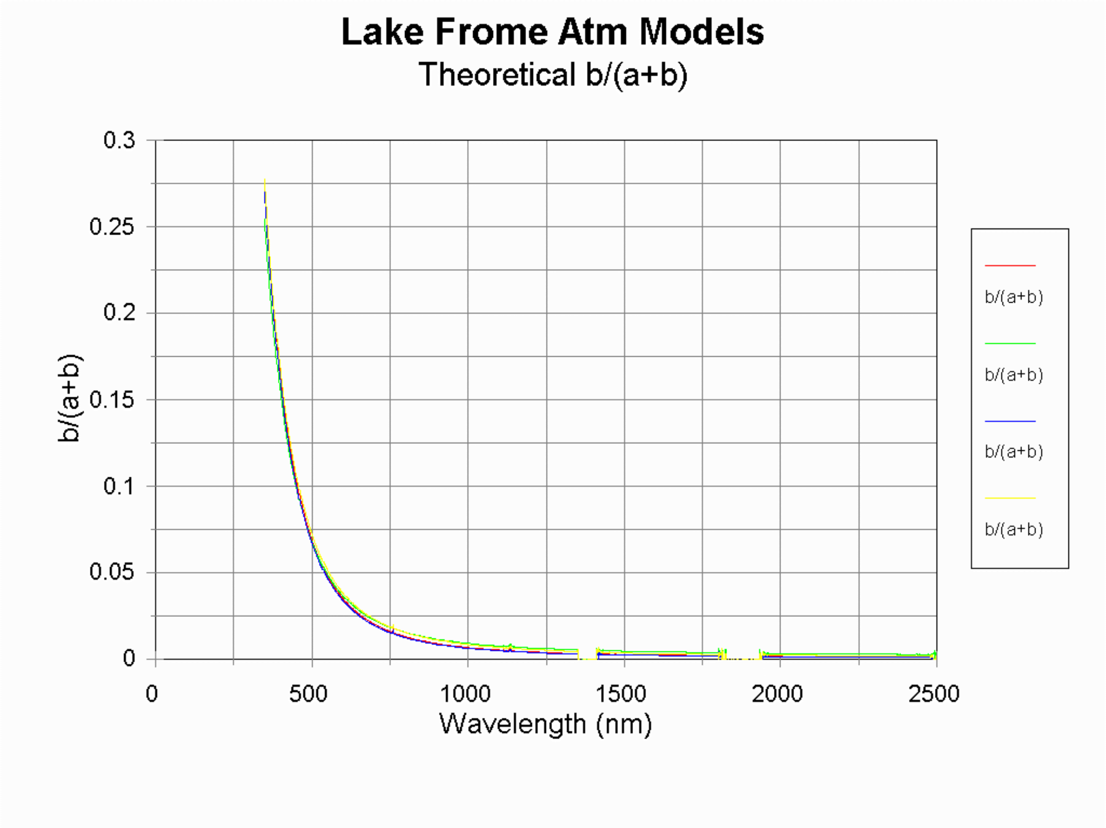
Note that the data do not even have to be calibrated to radiance to use this method and if there is a good range of spectral types (so that total $Var(\rho_j)$ is large at every wavelength) then maybe the spectra/PIFs can even be fair guesses. The extreme example is a pair of artificial targets – black and white. But the more spectra the better since having redundancy allows bad PIFs to be eliminated by using predictive error and jacknifing methods.

Assuming there are a number of PIFs and least squares is being used, you should *always* check for major differences between the PIFs ρ_{ij} and:

$$\tilde{\rho}_{ij} = \frac{L_{ij} - b_j}{a_j}$$

A model can be used to estimate *a* and *b* if atmospheric data are available. Calibrated Hyperion data can themselves be used to estimate some parameters (especially water vapour) and also to "tweak" the parameters to reduce the well-known spikes that arise in the model based approach. But further discussion of these issues can wait until another time.

As another check on the validity of the result you should form the ratio b_j / a_j or $b_j /(a_j + b_j)$ since experience and calculations show how these should be independent of the atmospheric absorption features and depend only on scattering to look somewhat similar to the following graph. It is also independent of calibration if the calibration factor needed is just a gain. Just how this can be built into a significant constraint on the simple model is under investigation.



Just to repeat, MNF and PCA with correlation matrix and most classifiers will be independent of this kind of simple atmospheric correction so that it is NOT necessary to do it before they are used. If multiple date images or spectral libraries are used, however, it is necessary to carry out some normalisation to approximate reflectance at target.

Geometry

Finally, the geometry of Hyperion is very nice. It clicks onto the map base with not much more than an affine or bilinear transformation.

A bilinear transformation of the form:

$$x' = a_{10} + a_{11}x + a_{12}y + a_{13}xy$$
$$y' = a_{20} + a_{21}x + a_{22}y + a_{23}xy$$

seems often to be optimum with regard to predictive error with the slight cross-effect due to pointing induced projective geometry which can vary between images.

I will repeat a Table used in the EOC/ANUTech IGARSS paper estimating the major image on-ground parameters based on an affine transformation (location, x and y stretch, rotation and skew):

TABLE I

HYPERION GEOMETRIC CHARACTERISTICS

Pixel-X	30.646 m
Pixel-Y	30.528 m
Rotation	-12.38°
Skew	0.006°

EO-1 has a specific yaw to compensate for the Earth rotation skew. It is done for radiometric performance of ALI rather than geometry. However, for the geometry it also seems to be very effective and useful. The pixel shape is also very close to square. The predictive errors for the VNIR sensor at Lake Frome were 15 metres in x and 20 metres in y.

The rotation is very close to the "effective heading" as obtained by the spreadsheet "Heading.xls" that is available to you in the "Utilities" section of the oz_pi area using equations described in "Heading.doc". This spreadsheet can therefore be confidently used to estimate the Hyperion track for field work.

There is, however, an on-ground shift in the view of the two arrays (VNIR and SWIR). It is about 1 pixel in magnitude. This is due to the read-out of the arrays (causing a skew between the VNIR and SWIR data) and a small across track misalignment. In areas where spectra change rapidly spatially this may give effects that are hard to interpret and apparent mismatch between the VNIR and SWIR. You should be aware of it and think about how to take it out possibly by resampling.

The findings on this and other geometric effects will be summarised by EOC people following the Frome exercise. A cross-matched set of two Landsat images and two Hyperion images separated into VNIR and SWIR was used at Frome for geometric testing. It is hoped to put at least one ALI image into this group. Mosmod is very useful here to simultaneously tie all the images together.

I believe you will all be able to register your data to a very reasonable level of accuracy with little problem. If the Landsat scene you order through the EOC is fully geo-rectified, the location of GCPs is very easy (but tedious) and location of GPS sites follows. It would be even easier and better with a correlation match GCP finder.

Documentation & Files

I hope we can build up documents and useful files on the oz_pi area as they are completed. Since it a "secure" area I am hopeful we can put the IGARSS special session papers there as well so that you can see what others are putting in.

Documents:

In the oz_pi area under "Media" there is a very good basic, but by no means shallow or compromised document entitled:

Hyp_3_00_MSSpaper_030100.doc

It is a good overview and you all should read it before tackling Hyperion data. It is not publicly releasable! That was a mistake.

Under the “More Documents” area are two TRW publications:

Hyperion_Cal.pdf
Hyperion_Performance.pdf

These also add some useful material to the mill and should be checked out for the extra information that may help you or explain what you are finding.

As mentioned above, under “Utilities” there is:

Heading.pdf
Heading.xls

These allow you to compute the effective heading of Hyperion on the ground for use in planning missions. It seems to work well for Frome anyway!

As noted in the text, FWHM_Notes.doc discusses the relationships between the fwhm in the cases of rectangular (casi), Hymap (triangular) and Hyperion (Gaussian) bandpass functions.

As more documentation appears (maybe from more reputable sources) it will be made available through the oz_pi area.

Files mentioned in the text to be found in oz_pi are:

Hyperion_Cen_fwhm.dat	calibration file
SpectralL0_RevA.dat	Central wavelengths for Hyperion
BandwidthL0.dat	FWHM for Hyperion
S047_Cal.dat	Current Level 1 calibration file