Optimal spectral band configuration for forest land-cover classification of hyperspectral data: a study for the Italian-Canadian Joint Hyperspectral Mission

Marco Gianinetto, Francesco Rota Nodari, Pieralberto Maianti, Riccardo Tortini and Giovanmaria Lechi

Politecnico di Milano, Laboratory of Remote Sensing (LaRS), Building Environment Sciences and Technology (BEST) Department, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
E-mail: marco.gianinetto@polimi.it

Abstract
In 2006 the Italian and the Canadian Space Agencies started a collaboration to evaluate the feasibility of the Joint Hyperspectral Mission (JHM), a new mission for Earth Observation devoted to environmental applications. JHM was designed to operate with a 30 m resolution hyperspectral sensor able to collect 210 narrow spectral bands in the range of 400-2500 nm. This paper presents a study carried on for the Italian Space Agency during Phase A, aimed to suggest an optimal spectral setup for the land-cover key application. Just referring to the mapping of forest species, results on simulated JHM data suggested that an optimal configuration can be obtained using a 50 nm bandwidth.

Keywords: Joint Hyperspectral Mission, hyperspectral data, land-cover, data simulation.

Introduction
Hyperspectral imaging is a fast growing area in remote sensing and improves the capability of multispectral image analysis. The main advantage of the hyperspectral imaging...
technology is that the image pixel generated generally contains more spectral information than does a multispectral image vector. From a theoretical point of view this should improve material detection, discrimination, classification and identification [Chang, 2003], but will also produce huge amount of data which makes their processing a serious computational challenge. Moreover, hyperspectral sensors typically use hundreds of contiguous spectral bands with very narrow bandwidth and this can produce low signal-to-noise ratio (SNR) during data acquisition at longer wavelengths. This is due to the small solar radiation available in the range of 2.0-2.5 μm, resulting in noisy image collection. Consequently, for land-cover classification the use of hundreds of contiguous narrow spectral bands in the optical domain may not be the best configuration and data reduction techniques together with methods for improving the SNR should be used to increase the computational efficiency and overall classification accuracy [Gianinetto and Lechi, 2004].

On June 2005 the Italian Space Agency (ASI) and the Canadian Space Agency (CSA) signed a memorandum of understanding concerning cooperation in the area of Earth Observation (EO) and in 2006 they started a collaboration to evaluate the feasibility of a future satellite mission, called Joint Hyperspectral Mission (JHM). The Italian-Canadian JHM looked for a small satellite (less than 800 kg) operating in a near circular sun-synchronous orbit at 700 km altitude and equipped with the following payload [Staenz, 2007; Ananasso et al., 2008]:

1) a 30-meter resolution hyperspectral camera in the range of 400-2500 nm;
2) a 5-meter resolution panchromatic camera in the range of 400-700 nm;
3) a 90-meter resolution medium infrared camera at 3.9 μm;
4) a 90-meter resolution multispectral thermal infrared camera in the range of 8.2-12.5 μm.

Regarding the hyperspectral payload, similarly to Hyperion, JHM was designed to collect 210 narrow contiguous spectral bands (10 nm bandwidth) with a SNR greater than 200:1 along the entire spectral region (peak SNR for typical target of 600:1 at 650 nm and 400:1 at 1750 nm) and a radiometric accuracy greater than 5% [Ananasso et al., 2008].

The JHM mission was focused on a wide range of users’ needs and key applications (e.g., coastal zone mapping, urban areas mapping and monitoring, water management, geological mapping, forest inventory, environmental mapping, land cover and land use, risk management and hazard monitoring) and was thought for meeting both scientific and operational objectives that were not achievable using existing EO technologies [eoPortal, 2008]. Moreover, with respect to risk management and hazard monitoring, JHM was thought as complimentary to the Italian COSMO-SkyMed constellation [Candela, 2008; D’Errico and Fasano, 2008], recently launched as part of the dual-purpose most sophisticated observation satellite network of Europe ORFEO.

From the Canadian side, JHM inherited the cooperation with the Australian Resource Information and Environmental Satellite (ARIES) mission [Ballinger, 2001], the cooperation with the German Aerospace Center on the Environmental Mapping and Analysis Program (EnMAP) [Stuffer et al., 2007] and the experience of the Canadian Hyperspectral Environment and Resource Observer (HERO) mission [Hollinger et al., 2006]. From the Italian side, JHM inherited the past experiences of the HyperSpectral Earth Observer (HypSEO) mission [Mochi et al., 2002], of the Mid Infrared Thermal High Resolution
Imaging System (MITHRIS) and of the Thermal High-resolution Earth MApper [Coppo et al., 2003].
This paper presents a study carried on for the Italian Space Agency during Phase A, aimed to suggest an optimal spectral setup for the land-cover key application.

**Methods**

**Spectral dataset**
This study was addressed to simulate the land-cover mapping capabilities of the future JHM system in forest and semi-natural areas (broad-leaved forest, coniferous forest and mixed forest).
As far as JHM was planned as a joint Italian-Canadian mission, the simulation considered a scenario of land-cover mapping of forest species consistent with both the Counties’ landscapes and also with the CORINE Land Cover. Six pure land-cover classes were defined (Aspen, Fir tree, Lodgepole pine, Maple, Oak and Willow) and high-resolution laboratory spectra (<10 nm) were used as source data [Clark et al., 2003]. Table 1 shows a summary of data and Figure 1 shows the source spectra used for the simulation.

<table>
<thead>
<tr>
<th>Scientific Name</th>
<th>Italian name</th>
<th>US name</th>
<th>Source spectral resolution (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Populous tremuloides</em></td>
<td>Pioppo tremulo</td>
<td>Aspen</td>
<td>8</td>
</tr>
<tr>
<td><em>Abies alba Mill</em></td>
<td>Abete bianco</td>
<td>Fir tree</td>
<td>8</td>
</tr>
<tr>
<td><em>Pinus contorta</em></td>
<td>Pino delle dune</td>
<td>Lodgepole Pine</td>
<td>1</td>
</tr>
<tr>
<td><em>Acer sp.</em></td>
<td>Acero</td>
<td>Maple</td>
<td>8</td>
</tr>
<tr>
<td><em>Quercus robur</em></td>
<td>Farnia</td>
<td>Oak</td>
<td>8</td>
</tr>
<tr>
<td><em>Salix sp.</em></td>
<td>Salice</td>
<td>Willow</td>
<td>8</td>
</tr>
</tbody>
</table>

**Simulation of the JHM’s data**
According to the purpose of the study, for each land-cover class at-sensor-radiance was simulated using the 6S radiative transfer code [Vermote et al., 1997] for a typical survey over central Italy, in summer time and with nadir looking geometry. The JHM’s simulated data were generated starting from the mission’s specifications, adding to the at-sensor-radiance computed the SNR and the radiometric accuracy as defined by the mission’s requirements. Table 2 shows a summary of the parameters used for the simulation.

To take into account for different setup of the hyperspectral payload, high-resolution input data were spectrally resampled (downsampled).
For each land-cover class ten synthetic reference signatures were generated in the range of 400 – 2500 nm, each of them with a different spectral resolution ranging from 10 nm bandwidth (210 spectral bands) to 100 nm bandwidth (21 spectral bands). Being a simulation, all the modeled bands were considered equally spaced and with a Gaussian
spectral response. Consequently, spectral resampling was performed assuming as critical sampling a Gaussian model with a Full Width at Half Maximum (FWHM) equal to the band spacings. Finally, the ‘natural’ variability of the classes was simulated: each of the six input signatures were resampled to the target desired spectral resolution (60 spectra) and each of the resampled signatures was added a random noise through a Monte Carlo simulation (40 independent extractions, for a total of 2400 simulated spectra).

Figure 1 - Source reference spectra from USGS Digital Spectral Library ‘splib05a’ [Clark et al., 2003] used in the simulation of JHM data. (a) Aspen; (b) Firtree; (c) Lodgepole Pine; (d) Maple; (e) Oak; (f) Willow.
Based on this dataset, ten synthetic radiometric data were produced: one for each target spectral resolution, containing both the source signatures (pure pixel) and mixtures in different percentage (mixed pixel). Mixtures were computed for 2-classes, 3-classes, 4-classes, 5-classes and 6-classes mixed pixel according to the following scheme:

\[ S_{JHM}(x,y) = \sum_{i=2}^{n} \alpha_i s_i \]  

where:

\( S_{JHM}(x,y) \) is the JHM simulated spectra for the generic pixel \((x,y)\);

\( s_i \) are the spectra of the pure classes;

\( \alpha_i \) are the percentage of mixtures of the pure classes \( \sum_{i=2}^{n} \alpha_i = 1 \).

Weights \( \alpha_i \) were selected to always retain a dominant species in the mixture (e.g. \( \alpha_1 = 0.30 \), \( \alpha_2 = 0.23 \), \( \alpha_3 = 0.23 \), \( \alpha_4 = 0.23 \) with \( 1 < i \neq j \neq k \neq m < 6 \) for a generic 4-classes simulated mixed pixels with \( i \)-th dominant species).

### Table 2 – Summary of the parameters used for the simulation with 6S.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Roma (Italy)</td>
</tr>
<tr>
<td>Date</td>
<td>25 July</td>
</tr>
<tr>
<td>Solar azimuth angle</td>
<td>117.01°</td>
</tr>
<tr>
<td>Solar zenith angle</td>
<td>36.38°</td>
</tr>
<tr>
<td>Viewing azimuth angle</td>
<td>0°</td>
</tr>
<tr>
<td>Viewing zenith angle</td>
<td>0°</td>
</tr>
<tr>
<td>Atmospheric conditions</td>
<td>Mid latitude summer</td>
</tr>
<tr>
<td>Aerosol model</td>
<td>Continental</td>
</tr>
<tr>
<td>Aerosol model concentration</td>
<td>Visibility=23 km</td>
</tr>
<tr>
<td>Target elevation</td>
<td>0 m (target at sea level)</td>
</tr>
<tr>
<td>Sensor altitude</td>
<td>700 km</td>
</tr>
<tr>
<td>SNR</td>
<td>200:1 (worst)</td>
</tr>
<tr>
<td>Radiometric accuracy</td>
<td>5% (on all the spectral region)</td>
</tr>
</tbody>
</table>

**Data classification and accuracy assessment**

In past, spectral unmixing algorithms have been often used for a quantitative analysis of sub-pixel spectral information (pixel mixtures). In particular, linear unmixing methods have been often preferred to nonlinear mixture models because simpler, more generic and proved to be useful in several remote sensing applications [Cross et al., 1991; Settle and Drake, 1993; Gong et al., 1994; Garcia-Haro et al., 1996]. In this study, the JHM’s simulated data were classified using the Linear Spectral Unmixing classifier (LSU). An overview about the use of LSU for hyperspectral data classification can be found in Li [2004].
For each spectral resolution 40 classification tests were performed using as training samples different spectral signatures randomly selected from the simulated data: one for each land-cover class, for a total of 400 data classifications. Finally, the classification errors were evaluated as a function of the spectral resolution and the optimal spectral configuration was determined. For each spectral resolution, classification errors were estimated for all the spectra of the synthetic radiometric data generated. Errors \( \varepsilon_i \) were calculated as the difference between the computed \( f_{c,i} \) and the real (known) class’ fraction \( f_{r,i} \), as follows:

\[
\varepsilon_i = f_{c,i} - f_{r,i} \quad [2]
\]

where:
\( \varepsilon_i \) is the classification error for the \( i \)-th spectra;
\( f_{c,i} \) is the computed class’ fraction for the \( i \)-th spectra;
\( f_{r,i} \) is the real (known) class’ fraction for the \( i \)-th spectra.

Each classification test had its own accuracy index and, consequently, a summary index for every land-cover class was determined. Under the hypothesis of Gaussian distribution it is possible to give, for every land-cover class, a mean classification error computed as the mean value of all the classification errors estimated (400 tests for each land-cover class). Moreover, fixed the spectral resolution and the land-cover class we had 40 values of classification accuracy (one for each simulated signature through the Monte Carlo approach). Under the hypothesis of Uniform distribution of errors the overall error and its acceptance region \( (\delta) \) for significance level \( \gamma \), can be computed using Equations [3] and [4], as follows:

\[
T(X) \cong E(X) = \frac{b + a}{2} \quad [3]
\]

\[
\delta = \left[ T(X) - \frac{T(X)}{\sqrt{1 - \gamma}} - T(X), T(X) + \frac{T(X)}{\sqrt{1 - \gamma}} - T(X) \right] \quad [4]
\]

where:
\( \delta \) is the acceptance region;
\( a \) is the minimum observed value;
\( b \) is the maximum observed value;
\( \gamma \) is the significance level.

**Results and discussion**
The hypothesis of Gaussian distribution for the classification errors \( (\varepsilon_i) \) is supported by the evidence that errors are independent variables (computed from independent pixel based data processing), thus their probability distribution function can be approximated with the Gauss error function \( (erf) \), as follows:
The latter assumption was verified using a Pearson’s Chi-square test ($\chi^2$) for significance level of 10%. For the data analyzed we obtained the following values:

\[
\begin{align*}
\chi_{obs}^2 &= 7.79 \\
F_{\chi_{0.05}^2 - 1 - \alpha}^{-1} (1 - \alpha) &= 24.77 \\
\alpha &= 0.1 \\
\end{align*}
\]

where:

- $\chi_{obs}^2$ is the observed Chi-square value;
- $F_{\chi_{0.05}^2 - 1 - \alpha}^{-1} (1 - \alpha)$ is the critical value for rejection;
- $\alpha$ is the significance level.

Being $\chi_{obs}^2 < F_{\chi_{0.05}^2 - 1 - \alpha}^{-1} (1 - \alpha)$, the null hypothesis of Gauss error function was accepted, and a mean classification error $m_{ij} = \bar{E}_i$ was computed for every land-cover class. As previously described, the 40 $m_{ij}$ values were assumed to belong to a Uniform distribution, therefore an overall index considered as representative of the overall classification accuracy was estimated using Equations [3] and [4]. The hypothesis of Uniform distribution for $m_{ij}$ was supported by the plotting of the observed vs. theoretical cumulative frequencies, as can be seen in the example of Figure 2 for Fir tree at 50 nm spectral resolution. The difference between the theoretical (plain line) and the observed values (dots) are due to a not ideal fitting of data. However, it is important to point out that the trend of the cumulative frequencies follows a Uniform distribution.

Figure 3 shows the classification error and its confidence region for all the land-cover classes here considered as a function of the spectral resolution. Generally speaking, the classification error increased with increasing the sampling interval for all the land-cover types. With the exception of Willow, that always showed a small classification error (<5%), errors ranged from about 2%-3% to about 15%-20%. All tests showed that at 50 nm sampling interval the classification accuracy increased, often reaching values similar to those obtained for the 10 nm sampling interval.

This simulation highlighted that for optical imaging sensor with contiguous spectral bands characterized by fixed bandwidth, the use 50 nm wide channels seems to be a good compromise between data volume, spectral richness, processing time and classification accuracy. Similar results suggesting the use of wider spectral bands instead of very narrow ones were obtained for MIVIS and Hyperion data processing for land-cover mapping of vegetated areas [Gianinetto and Lechi, 2004]. Moreover, it is important to note that when using very narrow spectral bands (e.g. with 10-20 nm bandwidth) the SWIR region is often very noisy [Gianinetto and Lechi, 2004]. Consequently, the use of 50 nm spectral bands should also involve the benefit of increasing the SNR according to wavelengths.
Figure 2 - Comparison of observed (dots) vs. theoretical (plain line) cumulative frequencies for Fir tree at 50 nm spectral resolution.

Figure 3 - Overall classification error computed from 400 classification tests for the following land-cover classes: (a) Aspen; (b) Fir tree; (c) Lodgepole Pine; (d) Maple; (e) Oak; (f) Willow.
Conclusions
This study evaluated for the Italian-Canadian Joint Hyperspectral Mission the classification accuracy expected for the land-cover key application. The simulation considered a scenario of land-cover mapping of forest species consistent with both the Italian and Canadian landscapes and with the CORINE Land Cover.
Results suggested that if contiguous spectral bands with fixed bandwidth are required, an optimal setup can be obtained using a 50 nm spectral resolution. This means that for the simulated data in the VNIR-SWIR domain (between 400 nm and 2500 nm), 42 spectral bands were enough to obtain similar accuracies than those obtained using the 210 narrower spectral bands scheduled for JHM. Moreover, the proposed configuration should be able to behave even better on real data due to the higher SNR in the SWIR region, as confirmed in other studies analyzing hyperspectral data of different nature. As an alternative, if contiguous spectral bands with variable bandwidth are allowed, the study suggested not to use bandwidth smaller than 50 nm for the SWIR region.

Acknowledgment
This research was carried on for the Italian Space Agency in the framework of the Italian-Canadian Joint Hyperspectral Mission and was supported only with internal funding. Authors are thankful to USGS for the availability of high-resolution spectra.

References
3373–3400.


Received 08/06/2009, accepted 18/09/2009.