Quality Assessment of CHRIS/PROBA Image and Recommendation for Land cover Classification

Niladri Gupta and E. J. Milton

School of Geography, University of Southampton, Southampton, SO17 1BJ, UK
Email: ng106@soton.ac.uk

Summary

Quality assessment of satellite data forms an important part for any land use classification process and as a result every user should know the quality of the image used for preparation of a map. In the present day scenario, data from small experimental satellites are being used a lot for preparation of maps for land and water resources application. To assess the quality of these products, accuracy is in general the true value of the quantity that is being measured. An accuracy statement enables the end user to have a first hand knowledge about the cost involvement and the algorithms to be used for data calibration. The present study involved the i) quality assessment of nadir view CHRIS data (raw and destriped) and ii) recommendation for mapping land cover units. The images were processed and destriped using 1) DIELMO destriping method and 2) ESSC destriping algorithm using HDF Clean. The study looks into the procedure followed by the above two methods of destriping the raw data and their outcome. But the question that arises whether these destriping processes are at all effective in generation of landuse / landcover map? A comparative qualitative assessment of the raw image and the two images destriped by two different methods, were carried out. A quantitative analysis of the three images were also undertaken to find out whether the enhancement of the images by way of destriping helps in the classification process? The analysis was undertaken by i) statistical approach and ii) classification of the three images using unsupervised classification technique. A comparative study of the classified images generated from the raw data and data cleaned by ESSC algorithm and DIELMO was also carried out. This study on the basis of the quantitative and qualitative analysis recommends the best images for landcover classification.

1 Introduction

Quality assessment of satellite data is an important part of any landuse classification process. According to Maling (1989) "Maps are deliberate generalization of reality" and as a result every user needs know the quality of the images utilized for preparation of the maps. Accuracy is in general the true value of the quantity that is being measured. A complete accuracy statement enables the end user to get a first hand knowledge about the cost involvement and the algorithms to be used for preparation of a land use / land cover map.

The present study involves the following

i) Quality assessment of CHRIS data (raw and destriped) of Chichester Harbour in South western coast of England.
ii) Recommendations for a land cover classification process.

1.1 Assessment of Raw Data

CHRIS is an imaging spectrometer, a new generation of satellite sensor with resolution in between air borne and space borne imaging spectroscopy. This new generation sensor is affected mainly by two noises a) horizontal noise and b) vertical noise. The horizontal noise is manifested by partial loss of data in different band position while the vertical noise or striping is caused either by sensor alignment or due to thermal fluctuation of optical elements during satellite orbit (Garcia and Moreno, 2004). The raw data of the Chichester Harbour area is affected by the vertical striping as can be seen in the Figure 1(a).

But it has been observed that if individual bands of the original image are taken into consideration the striping phenomenon is more prevalent in the lower bands than the higher bands (Gracia and Moreno, 2004) as can be seen in Figure 1(b) and Figure 1(c) respectively.

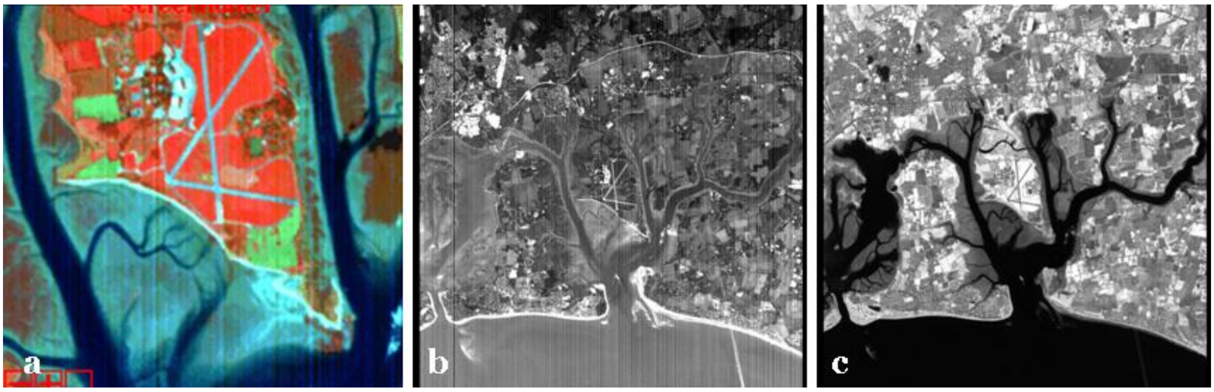

Figure 1. (a) Vertical Striping in CHRIS Data (FCC, Bands 12, 8, 1), (b) CHRIS Image (Band 1) and (c) CHRIS Image (Band 12).

In this case study two destriping algorithms are used to see how it affects the classification process. The two methods being used are i) DIELMO method (Garcia and Moreno, 2004) and ii) HDF Clean developed by J. Settle and Surrey Satellite Technology Limited, and distributed by ESA.

The DIELMO method demonstrates that the vertical striping can be cleaned by using a correction factor without use of any filter and moreover as there is no dependence on the image content while correcting the error so the same factor can be applied on all the images from 5 different angles. But they also point out that the high frequency vertical noise can be corrected easily but not the low frequency ones (Gracia and Moreno, 2004).

The other process uses software developed by Dr. Jeff Settle, University of Reading, United Kingdom which readjusts the relative brightness of pixels after assessing the average pixel value within an inter-quartile range.

2 Study Area

Now the question arises whether these destriping processes at all affect the generation of landuse / landcover map of the mud flat area? Taking this into account the first task would be to look into various landuse categories that are present in this area of interest. This may be looked into as discrete scene model in which every feature is characterized by a particular property and similar features have the same property. In addition to that there may be some background elements which are spatially continuous, having uniform property but partly obscured by other scene elements (Strahler *et al.*, 1986). By looking at the imagery (both striped and destriped) as individual bands and as composites (FCC and true color) it can be said that the image is a nested discrete model in which individual elements can be identified and some elements aggregate together to give another element e.g. the eel grass or may be the algae. Though definitely it can be said that the some landuse categories are identifiable individually as well as in a group, the identification of algae is matter of question where the resolution of image comes into play. As per technical specification of CHRIS sensor it can operate in two modes (i) Full SWATH, reduced spatial resolution and (ii) Full SWATH, full spatial resolution. The data supplied for usage in the project has been acquired in the second mode as it has a high spatial resolution (17m) over full SWATH and in 18 spectral channels (Barnsley et al, 2004).

The image of the area of interest has a very high spatial resolution and as a result can be considered to be an H-resolution model where individual entities can be identified. Moreover the background information of Chichester Harbour (Figure 2) reveals that the major components are eelgrass, marine algae and spring crops which are major food source of the internationally important population of wading birds, specially the dark-bellied Brent Geese (*Branta bernicla*). These elements can be considered as the scene elements and are larger than the spatial resolution of the pixel and does not require parameterization as they can be identified separately. It is also possible to establish a direct relationship between class of scene elements and measurement from a sensor i.e. a simple, empirical and invertible remote sensing model (Strahler *et al.,* 1986). Besides the above classes a few other scene elements can also be identified from the data of the area of interest e.g. the water channel area where the river meets the sea, the exposed mud flat, pebbles, runway and some settlement areas.

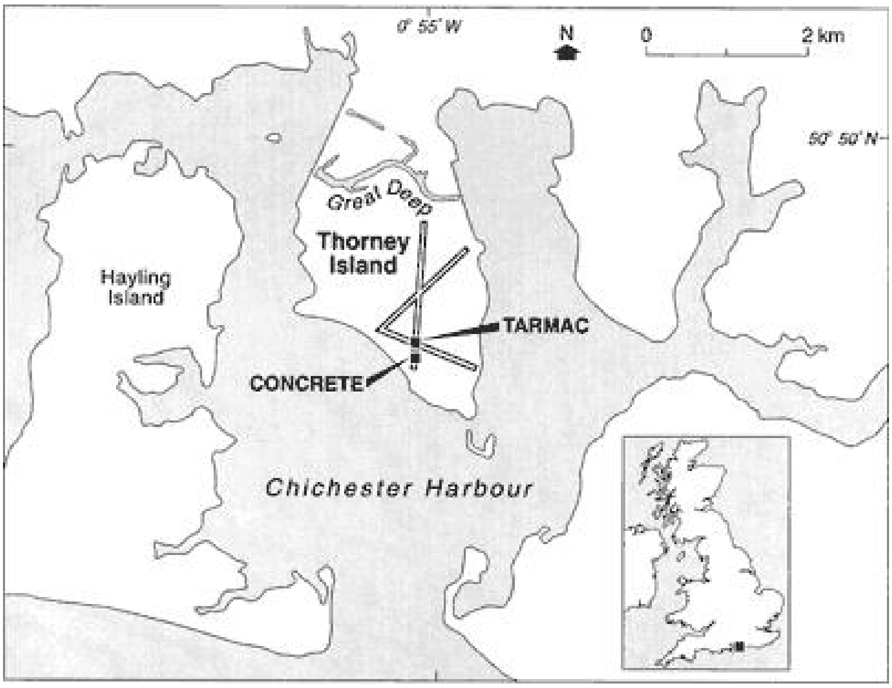

Figure 2. The location of the Chichester Harbour, the CHRIS PROBA test site (Milton et. al, 2006).

But according to various studies done so far the eelgrass and the algae is most of the time associated with each other and to identify them separately, in situ spectral radiance measurement required to be taken in the field and then compared with the satellite data (generally hyperspectral data) for their zonation (Albertoanza *et al.,* 1999). The aerial photo provided also does not give any indication of the distinction of eelgrass and the algae. So in this situation the scene model can be a mixture of H – resolution and nested discrete model or an L-resolution model.

3 Data Assessment

3.1 Qualitative Analysis

Considering the above factors of individual scene elements a comparison of the three individual images can be done. As far as landuse mapping is considered where vegetation cover is one of the most prevalent category a combination of green, red and near infrared band is the most suitable one. Green has the maximum reflectance in the NIR wavelength bands. Another important component of the area is the mudflat and the river channels. These components can also be deciphered using the above FCC. Moreover, the air photo also throws some light on the landuse categories present in the area though one of them is about 10 years old. The ENVI software has been used to make a comparative qualitative analysis of the three images and also with the aerial photo. Besides false colour composites, true colour composites were also prepared for the three images for comparing with the aerial photo. Though the aerial photo and the 3 images were of two different season spring and autumn respectively, but both these periods are the time for abundance of algae and eel grass. In addition the images were of a low tide time so the mudflat is exposed which actually helps in delineating the landuse / landcover boundaries. But visually the destriped images look much better than the striped ones as seen in Figure 3 (a), (b) and (c).

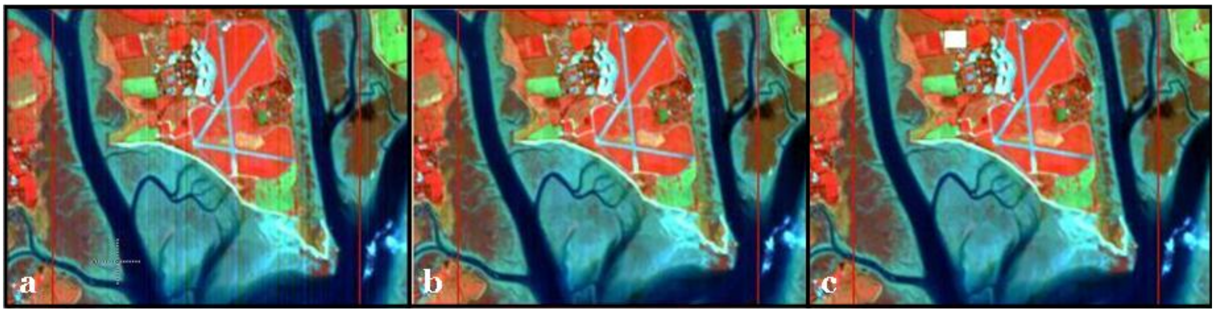

Figure3 (a) Striped FCC data, (b) FCC data cleaned by HDF Clean (c) FCC data cleaned by DIELMO Method.

But if we take into consideration the HDF clean and the DIELMO Data as seen in Figure 3(b) & (c) there is no visually identifiable difference between the two images, so in that case the data destriped by the HDF Clean method can be preferred over the image destriped by the DIELMO Method.

3.2 Quantitative Analysis

A quantitative analysis of the three CHRIS data is considered to find out if the destriping process helps in enhancement of the image and in the classification process. This can be done either by taking a statistical approach or by classification of the three images. The statistical approach involves taking the mean median and standard deviation of pixel values of various identifiable categories in all the three images and studying the difference between them. This was done by taking ROIs of spectrally different objects present in all the images. A statistics was then generated for each of the objects in all the 3 images. The standard deviation graphs look more or less similar except a notch at the 700 nm band in case of the raw images which is absent in the two destriped image. This is due to the reduction of signal noise ratio caused by the striping in the images. But if we look at the mean and standard deviation of the bands that are more useful for classification of image (12,8,1) it can be seen that there is not variation in mean values of the raw data and the HDF clean data while the difference with DIELMO one is quite considerable especially in the bands 12 and 8, but the difference in band 1 is quite high as it is the one which is most affected by the striping phenomenon as a result of which it is the most modified. Similarly, if we look into other ROIs it also reveals that above 700 nm wavelength i.e. in NIR band some modification can be identified as seen in Figure 4 (a), (b) and (c). But if we look at the statistics the difference of mean and standard deviation is more pronounced in the lower bands and in addition it is more prominent in the DIELMO data. Table 1 shows the statistical difference in three regions of the three different images. So, the above statistical analysis shows that there is not much difference in statistics of the image except that visually there is no striping in the cleaned image.

Table 1. Comparative Statistics of Spectrum (Band 12, 8, 1: FCC) in Region 1 , 2 and 3 of satellite data.

Band	**Region # 1**		**Region # 2**		**Region # 3**	
	Mean	**Std. Deviation**	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
Raw Data						
1	49177.035	819.949	46490.488	978.280	45834.960	646.754
8	20258.640	946.527	15098.760	1349.397	9104.780	274.155
12	27812.687	722.650	33238.385	2476.816	5398.198	106.914
HDF Clean Data						
1	49238.756	596.441	46338.050	842.286	45781.872	429.490
8	20323.407	1015.012	15049.195	1362.534	9089.547	313.556
12	27572.430	725.171	33173.450	2504.768	5409.233	92.814
DIELMO Data						
1	49298.407	609.171	46491.414	827.923	45804.795	476.989
8	20370.360	950.949	15115.592	1395.618	9100.924	243.165
12	28324.035	869.140	33197.056	2417.559	5357.657	163.731

Another process of quantitative analysis of the images can be by operating classification technique on the images. An unsupervised classification was applied on all the three images with about 10 classes and 5 iteration using the K-mean method as the algorithm is more robust. The classified raw image shows the effect of striping while the other classified maps are same as seen in Fig. 5 (a), (b) and (c).

Figure 5 Unsupervised classification using 10 classes and 5 iteration using K-mean method performed on (a) Striped FCC data, (b) FCC data cleaned by HDF Clean (c) FCC data cleaned by DIELMO Method.

Even when a comparison is done between the two sets of data i.e. Raw and HDF Clean and Raw and DIELMO the change matrix shows a similarity of above 90% in most of the classes generated by unsupervised classification as can be seen in Table 2.

Table 2. Change detection statistics (similarity %) of striped and destriped data considering 10 classes.

Unsupervised Classes	**Raw and HDF Clean Data**	**Raw and DIELMO Data**
Class 1	99.24	99.27
Class 2	96.48	95.72
Class 3	88.63	91.37
Class 4	95.92	93.30
Class 5	91.28	82.46
Class 6	93.10	86.54
Class 7	94.71	84.17
Class 8	92.36	90.27
Class 9	94.50	90.95
Class 10	96.99	96.97

But as far as background information is concerned the intertidal mud flat and surroundings can have following categories of landuse:

i) eelgrass, ii) exposed mud flat zone iii) water channel area iv) pebbles v) agricultural land vi) settlement area and vii) algae. But the algal mass may be mixed with the eelgrass and may not always be identifiable from the satellite data. In that case there is no requirement for doing an unsupervised classification using ten classes. So, if we undertake an unsupervised classification using seven or 8 classes there is no difference between the classified images as seen in Fig. 6 (a), (b) and (c).

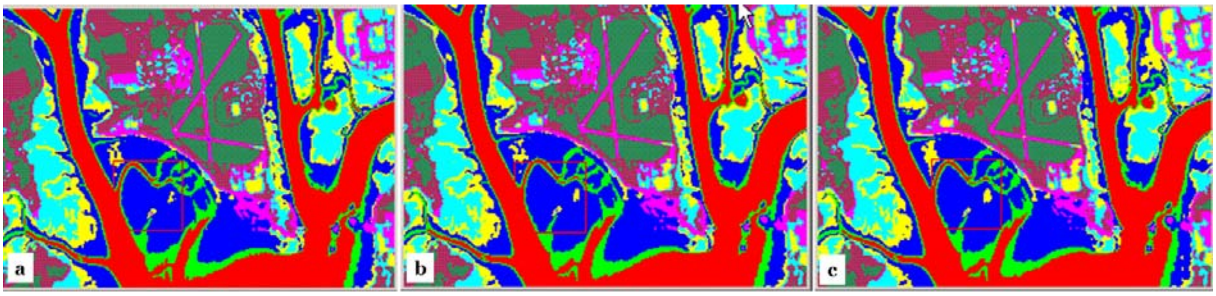

Figure 6 Unsupervised classification using 8 classes and 5 iteration using K-mean method performed on (a) Striped FCC data, (b) FCC data cleaned by HDF Clean (c) FCC data cleaned by DIELMO Method.

The above table indicates that if we consider 8 classes the change detection statistics show a similarity of above 96% when comparing between Raw and HDF Clean data while a similarity of above 94% when comparison is done between Raw and DIELMO data as cab be seen in Table 3.

Table 3. Change detection statistics (similarity %) of striped and destriped data considering 8 classes.

Unsupervised Classes	Raw and HDF Clean Data	Raw and DIELMO Data
Class 1	99.35	99.25
Class 2	96.75	95.54
Class 3	97.89	97.20
Class 4	93.93	92.27
Class 5	93.89	89.69
Class 6	94.26	93.15
Class 7	95.71	92.66
Class 8	97.48	96.97

4 Recommendations and Suggestions

Considering the above qualitative and quantitative aspects it can be inferred that there is no requirement for using the DIELMO destriped data as the destriping has no significant effect on the image statistics and post classification results. Regarding using the HDF Clean data it can be suggested that if the number of classes in the landuse map exceeds ten then only the effect of vertical stripes is identifiable in the image but it is not if the number of classes are eight or less. So in this case the CHRIS raw image can be used successfully used for preparation of landuse / landcover map as the number of classes won't exceed eight.

Regarding processing technique of the satellite data it is better to start with unsupervised classification technique as the area is unknown to us but a reconnaissance survey is required with multi-spectral field radiometer of the mudflat area especially to identify the algal mass and then map them by matching the spectral similarity (Alberotanza *et al.,* 1999). Then a post field supervised classification is recommended to take care of any misinterpretation in the unsupervised classes. In doing so if there is any increase in number of classes (> 10) we can then use the HDF Clean Data as there is no financial implication on using that data. Moreover, the DIELMO method has improved since this work was undertaken.

5 Acknowledgements

The author would like to express his gratitude to the European Union and the Erasmus Mundus Consortium for awarding the scholarship and this paper is an output of the GEM course undertaken by the author under this program.

References

ALBEROTANZA, L. B., BRANDO, V. E., RAVAGNAN, G. and ZANDONELLA, A., 1999. Hyperspectral aerial images. A valuable tool for submerged vegetation recognition in the Orbetello Lagoons, Italy. *International Journal of Remote Sensing,* **20**:523-533.

BARNSLEY M.J., SETTLE J.J., CUTTER M.A., LOBB D.R., and TESTON F., 2004. The PROBA/CHRIS Mission: A Low-Cost Smallsat for Hyperspectral Multiangle Observations of the Earth Surface and Atmosphere. IEEE Transaction of Geoscience and Remote Sensing, **42**(7): 1512 – 1520.

GARCIA, J.C., and MORENO, J., 2004. Removal of noises in CHRIS/PROBA images: Application to the SPARC campaign data. *Proceedings of the 2nd CHRIS/PROBA Workshop,* ESA/ESRIN, Frascati, Italy; 28-30 April, 2004.

MILTON, E. J., GUANTER, L., GARCÍA, J. and ANDERSON, E., 2006. The contribution of CHRIS PROBA to NCAVEO, a knowledge transfer network for the calibration and validation of Earth observation data. *Proceedings of the 4th CHRIS/PROBA Workshop*. http://earth.esa.int/workshops/4th_chris_proba/, 19-21 September 2006.

MALING, D. H., 1989. Measurements from Maps: Principles and Methods of Cartometry. (Oxford: Pergamon Press).

STRAHLER, A.H., WOODCOCK, C.E., and SMITH, J.A., 1986. On nature of models in remote sensing. *Remote Sensing of Environment*, **20**: 121 – 139.