MERIS Cloud Masks: Exploration and Visualisation of MERIS Spectra

Dr Dan Cornford\textsuperscript{(1)}, Dr Anton Lyaskovskiy\textsuperscript{(1)}

\textsuperscript{(1)}Aston University, Aston Triangle, Birmingham B4 7ET, United Kingdom, Email: d.cornford@aston.ac.uk

ABSTRACT

In this work we present a cloud mask for MERIS, developed as part of the EU NAOC project, which can discriminate between optically thick and thin clouds. The method is based on the expert selection of a small labelled data set of cloudy and cloud free pixels in MERIS observations taken over the ocean, guided by meteorological knowledge. This small labelled data set is augmented by a larger unlabelled data set randomly extracted from a number of MERIS scenes over the ocean. This unlabelled data set is used to characterise the structure of the MERIS spectra that are observed, using pattern recognition methods called the generative topographic mapping and Neuroscale. The generative topographic mapping constructs a density model for the 16 dimensional (i.e. the MERIS bands and a ratio between the radiances at the 11\textsuperscript{th} and the 10\textsuperscript{th} bands) data in a lower (typically 2) dimensional latent space, which allows visualisation and understanding of the structure and distribution of the data. The Neuroscale algorithm is a distance preserving data projection algorithm without a density model. The lower dimensional structure is then used to define a non-linear projection, which retains information, but permits the construction of simpler classification models, something that will be especially important with future hyper-spectral instruments. We show the results of our cloud classification on several MERIS scenes and contrast our cloud mask with the standard MERIS cloud mask.

1 INTRODUCTION

The traditional applications of neural networks for ocean colour use radiance or reflectance spectra as inputs \cite{1, 2}. The classification of remote sensed imagery is usually supervised. One of the problems with cloud and water type classification is the difficulty of obtaining a great number of labelled samples for the training dataset which represent the diversity of all possible situations, however unlabelled data are available in abundance.

Unlike supervised classification, clustering methods (or unsupervised methods) require no labeled training sets at all. Instead they attempt to find the underlying structure automatically by organizing the data into classes sharing similar spectral characteristics. Cluster analysis provides a useful method for organizing a large set of data so that the retrieval of information may be made more efficiently. The popular clustering algorithms used in remote sensing are the k-means clustering, agglomerative hierarchical clustering (AHC) \cite{3} and Probabilistic Self Organizing Map (PRSOM) \cite{4, 5}. The k-means algorithm is a method for finding \(K\) vectors that represent an entire dataset. The data is considered to be partitioned into \(K\) clusters, with each cluster represented by its mean vector and each data point assigned to the cluster with the closest vector. The AHC algorithm begins by assigning each pixel in the dataset to its own class or cluster. The next step is to find the two closest clusters according to some dissimilarity measure and agglomerate them to form a new cluster. The subsequent iterations require to utilize a dissimilarity between the merged clusters and all other clusters. Neither k-means nor AHC algorithms form a density model while the PR SOM algorithm approximates the probability density function of the data with a mixture of normal distributions. At the end of the learning phase, the probability density of the data estimated by the PR SOM map can be used to provide a classification.

In this work, we applied the Generative Topographic Mapping (GTM) \cite{6} and the Neuroscale \cite{7} algorithms to large volumes of unlabelled data. The GTM algorithm allows a non-linear transformation from latent space to data space with a fully probabilistic, generative model. In this approach, the data are modelled by a mixture of Gaussians, in which the centres of the Gaussians are constrained to lie on a lower dimensional manifold. Neuroscale is a non-linear distance preserving projection algorithm, similar in spirit to Sammon mapping. These models were then used to define a lower dimensional manifold on which the observations exist. Subsequently a smaller amount of labelled data was used to define the model that provides the classification on this lower dimensional subspace. In this work the labelling was obtained using meteorological expertise, but this could readily be augmented with physically based model data as models of cloud radiative properties improve. The classification was contrasted using both reflectances, and their projections on the manifold, as inputs. Both approaches were tested on a number of MERIS images and the results were
compared with the classification produced by a Multi-layer Perceptron (MLP) from reflectance spectra as well as the MERIS cloud mask.

2 ALGORITHMS

We apply two projection methods, GTM and Neuroscale, to the MERIS observations, which we briefly review below. In the GTM algorithm the projection of the data points from the latent space to the data space is defined by a Radial Basis Function (RBF) network [6]. In the GTM algorithm the d-dimensional data \( x = (x_1, \ldots, x_d) \) are represented in a q-dimensional latent variable space \( z = (z_1, \ldots, z_q) \). The two spaces are linked by an RBF, \( f(z,W) \), parameterised with the matrix \( W \), which maps the q-dimensional latent space, \( z \), to a manifold \( S \), embedded in the data space.

For this model to be useful, it is necessary \( q<d \). In fact GTM is most practical when \( q = 1 \) or 2. If we define a probability density \( p(z) \) on the latent space, this will induce a density \( p(y|W) \) in the data space. Since \( q<d \), this density will be zero away from the manifold \( S \). This is an unrealistic constraint, since we cannot reasonably expect the data to lie exactly on a q-dimensional manifold. Hence we add a noise model for \( x \) where it is convenient to use a spherical Gaussian with variance \( \sigma^2 \), so that the data density conditional on the latent variables is given by

\[
p(x|z,W,\sigma) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left\{ -\frac{\|f(z;W) - x\|^2}{2\sigma^2} \right\}
\]

The density in data space is then obtained by integrating out the latent variables:

\[
p(x|W,\sigma) = \int p(x|z,W,\sigma)p(z)dz
\]

Let the density \( p(z) \) be given by a sum of delta functions centres on nodes \( z_1, \ldots, z_M \) in latent space:

\[
p(z) = \frac{1}{M} \sum_{j=1}^{M} \delta(z - z_j)
\]

If the nodes are uniformly spread in latent space, this is an approximation to a uniform distribution. Equation (3) is now tractable and becomes a simple sum of \( M \) Gaussians:

\[
p(x|W,\sigma) = \frac{1}{M} \sum_{j=1}^{M} p(x|z_j,W,\sigma)
\]

This is a mixture model where all the kernels have the same mixing coefficient \( \frac{1}{M} \) and variance \( \sigma^2 \), and the jth centre is given by \( f(z_j;W) \). It is a constrained mixture model because centres are not independent but are related by the mapping \( f \). If this mapping is smooth, then the centres will necessarily be topographically related in the sense that two points \( z_a \) and \( z_b \) which are close in the latent space will be mapped to points \( f(z_a;W) \) and \( f(z_b;W) \) which are close in the data space.

The Neuroscale algorithm is a different approach based not on a density model, but instead on the concept of data topology. This is assumed to be captured by the inter-point distances, usually measured with a Euclidean metric, \( d^{*}_{ij} = \|x_i - x_j\| \). Each data point \( x_i \) is projected by a RBF to a point \( y_i \) in the lower dimensional latent space. The distance between points \( y_i \) and \( y_j \) is denoted by \( d_{ij} \) and the quality of the projection is measured by the Sammon stress metric:

\[
E_{sam} = \sum_{i=1}^{N} \sum_{j>i}^{N} \left( d_{ij} - d^{*}_{ij} \right)^2
\]
The smaller the stress, the more closely the distances between the $y_j$ match the distances in the original data space between the $x_i$ and hence the better preserved is the data structure in the projected space used for visualisation.

For the classification problem a feed-forward MLP is used in which we have a set of mutually exclusive classes (i.e. “thick cloud”, “thin cloud” and “open water”), and thus use the softmax activation function of the form

$$c_k = \frac{\exp(a_k)}{\sum_k \exp(a_k)}$$

(6)

where $a_k$ are the output layer activation values and $c_k$ are the output values of the network, which represent the conditional probability for class $k$ given the inputs presented to the network [6].

3 RESULTS AND DISCUSSION

For the unlabelled dataset we randomly extracted four thousand samples of the top-of-atmosphere (TOA) reflectances at 15 MERIS bands corrected for the “smile” effect by BEAM software (http://envisat.esa.int/services/beam/) from reduced resolution MERIS images. In addition to the band reflectances we also included a ratio between radiances at the 11th and the 10th bands in order to improve discrimination between the “thin cloud” class and the “thick cloud” class by taking into account atmospheric oxygen absorption and hence the optical path length. The BEAM toolbox was used to generate a dataset with samples labelled by an expert according to their meteorological and spatial context as well as their reflectance spectra.

Visualisations produced by GTM (Fig. 1a, b) and the Neuroscale (Fig. 2a, b) algorithms show that there is a lot of structure in both labelled and unlabelled datasets. The labelled data visualisation in both cases show good separation of classes. At the same time there are some areas in the unlabelled data visualization plots where we do not have sufficient labelled data, and an interesting extension of this work would be to identify those regions and explore their physical significance.

The classification results produced by MLPs trained on spectral TOA reflectance data with the oxygen absorption ratio (16 inputs) and trained on the manifold projection coordinates (2 inputs) produced by the Neuroscale algorithm learned on the unlabelled data are shown for a region of the North sea for two dates: 22nd September, 2003 (Fig. 3) and 25th September, 2003 (Fig. 4). They are compared with the corresponding TOA reflectance at 753.75 nm images with enhanced contrast which shows the thin cirrus clouds more clearly and the cloud masks from the corresponding Level 2 products. Both neural networks show similar results and produce more conservative cloud masks especially in the case of “thin cloud”. Large numbers of pixels classified as “thin cloud” by the neural networks are marked as “water” on the Level 2 images and consequently they were used for chlorophyll concentrations retrievals (Fig. 4b, c, d). This produces features in the chlorophyll retrieval, which are clearly due to the effects of the unidentified thin cirrus. Analysis of the spatial context of the features classified as “thin cloud” in the highlighted area over the North Sea west of the Dutch coast on the image for 25.09.2003 (Fig. 4c, d) strongly suggests that the detected features are due to thin clouds and not to water-leaving reflectance since they can be identified as part of a cloud system that extends over the land. Also the values of the oxygen absorption ratio for these pixels are considerably higher than those observed over the open water.

The described approach can be considered as a preprocessing procedure used for data exploration before retrievals of atmospheric and oceanic products. The use of the projective methods with multi-spectral data provides a useful visual representation of the structure of the observed reflectances, which can be used to explore the real instrumental data and its relation with our expectations from physical understanding. Where possible optical models for cloud should also be incorporated into the labelling procedure, but at present we feel that the use of meteorological expertise and spatial and temporal context provides a better characterization of the observed data. The probabilistic models like GTM can be also used for novelty detection – that is the determination of those observations made by a satellite which are unusual with respect to the data contained in the training datasets.
4 ACKNOWLEDGMENTS

The research was funded within the Neural network Algorithms for Ocean Colour (NAOC) project, a European Union Framework V programme (EVG1-CT-2000-00034).

5 REFERENCES

Fig. 1. MERIS data projection produced by GTM. (a) Unlabelled samples. (b) Labelled samples: red dots are “thick cloud” class, yellow dots are “thin cloud” class, blue dots are “open water” class

Fig. 2. MERIS data projection produced by Neuroscale. (a) Unlabelled samples. (b) Labelled samples: red dots are “thick cloud” class, yellow dots are “thin cloud” class, blue dots are “open water” class
Fig. 3. Classification results of the North Sea RR MERIS image, 22.09.2003. (a) TOA reflectance at 753.75 nm, (b) MERIS Level 2 product cloud mask, (c) cloud type classification by MLP trained on reflectance data, (d) cloud type classification by MLP trained on Neuroscale manifold projections.

Fig. 4. Classification results of the North Sea RR MERIS image, 25.09.2003. (a) TOA reflectance at 753.75 nm, (b) MERIS Level 2 product cloud mask, (c) cloud type classification by MLP trained on reflectance data, (d) cloud type classification by MLP trained on Neuroscale manifold projections.