Forecasting Research

Forecasting Research Division
Technical Report No. 206

Satellite Imagery Pattern Recognition Products for Numerical Weather Prediction

by

George Pankiewicz

January 1997

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

Satellite Imagery Pattern Recognition Products for Numerical Weather Prediction

NWP Division Technical Report No. 206

GEORGE PANKIEWICZ

NWP Division, Meteorological Office, London Road, Bracknell, Berkshire, RG12 2SZ, UK

Contents

Abstract

Following a number of pilot studies which looked at the potential for the use of satellite imagery pattern recognition in numerical weather prediction (NWP), a list of ten candidate products covering a wide variety of observation types has been drawn up. These include seven products aimed at improving current NWP observations, namely cloud cover, cloud type, synthetic relative humidity profiles, precipitation rate, radar anaprop removal, satellite winds and sea-ice extent. New NWP observations have also been considered, and include cloud system recognition for improving forecast error, objective PAOBs (paid observations) and potential vorticity mapping using water vapour imagery. This technical report describes each of the products in terms of their possible benefit to NWP as well as how they might be constructed and tested, considers likely implementation in UK Met. Office models and gives an estimate of the expected resources.

1 Introduction

Over the last three years, a project has been undertaken in the Satellite Image Applications Group of NWP(SA)[1], to examine the potential for new meteorological satellite imagery products obtained using a pattern recognition approach. Initially, the emphasis was on the identification of both cloud fields (such as cirrus and cumulus) and synoptic-scale cloud systems (cold fronts, warm sectors and depressions, for example), but recent work has pointed the way to developing many other products that in some way make use of the correlation between the spatial information contained in satellite imagery and other meteorological parameters beside cloud amount and type.

During this period, a number of tools were developed and studies carried out to investigate the pattern recognition approach, and to see whether it might be of benefit to NWP at the UK Met. Office (UKMO). These included:

1. A literature review of past and current cloud pattern recognition techniques and their application in meteorology, published in Meteorological Applications (Pankiewicz, 1995).

2. A Feature Selection and Evaluation Tool (FSET), built in conjunction with EDS Ltd (UK) to determine the optimal set of pattern recognition features for use in a classifier, given either objective or subjective training data.

   If a number of characteristics of an image are calculated for different classes of interest (for example, maximum infrared brightness temperature and visible standard deviation in a small region, in order to distinguish cumulus cloud from

[1]Numerical Weather Prediction (Satellite Applications)

water), then feature selection is the process of determining which characteristics are the most significant at discriminating the classes. In this example, visible standard deviation is expected to be the best discriminator, since for a field of cumulus, the maximum infrared brightness temperature might be exactly the same as that for a sample of water. In practice, probabilistic distance measures are used, such as the Fisher and Bhattacharyya distances. (Note that in pattern recognition literature, features are also often referred to as cues.)

3. A versatile MLP[2] neural network pattern classifier, coded in the PV-WAVE[3] image processing language, which can be regarded as an essential building block for any pattern recognition scheme.

   Such a classifier is referred to as a *supervised* classifier, since a set of training samples, each labelled either subjectively (with manually identified cloud fields, for example) or objectively (such as rainfall rates from radar data), are used to train the network into learning the classes (from the sample labels) by giving it the relevant features or cues.

   There are many good reasons for using neural networks as classifiers as opposed to more conventional methods such as Gaussian maximum likelihood and thresholding. These include the ability of neural networks to solve non-linearly separable problems and training to be data driven (removing the need to make assumptions about class distributions). The MLP has been coded to allow the user to make any number of modifications during training and classification, resulting in a network that can be tuned for use in many different applications.

4. A neural network airmass classifier, trained on 2190 semi-subjectively chosen Meteosat visible and infrared image pairs taken throughout 1994 to an accuracy of 80.8±2.2% (tested on 730 independent samples).

   The network classifies imagery into clear land or sea as well as three simple "airmass" classes, each associated with different cloud fields, namely dynamic, shallow convection and deep convection. The classifier was used successfully between May and October 1996 as part of a trial operational version of the GANDOLF[4] thunderstorm flood forecasting system, being built at the UKMO for the UK Environment Agency (Collier *et al.*, 1995; Pierce *et al.*, 1995; Pankiewicz, 1997a).

5. A two component neural network airmass classifier (following on from the GANDOLF classifier), using class dependent textural information provided by the original airmass classifier in a second, larger scale network.

   The aim of such a network is to improve the definition of the boundaries between airmasses, enabling a segmentation of airmass regions with a view to identifying synoptic-scale cloud structures (Pankiewicz, 1996).

---

[2]MultiLayer Perceptron

[3]Precision Visuals - Workstation Analysis & Visualisation Environment

[4]Generating Advanced Nowcasts for Deployment in Operational Land-based Flood forecasts

6. An investigation into the use of feature selection for satellite rainfall estimation in mid-latitudes, using the EDS FSET (Batten, 1996; Pankiewicz, 1997b).

   The best features were obtained from collocated radar and polar-orbiting NOAA AVHRR[5] imagery during spring 1991 in northwest Europe (the GPCP AIP-2[6] case study - WMO, 1994), and were composed of 8 spectral and textural features from AVHRR channels 3 and 5.

   A semi-supervised neural network built by EDS showed average classification accuracies for four rainfall rate bands, with separating thresholds of 0.125, 1.00 and 8.00 mmh^{-1}, of 55.7±5.3%. At the no-rain/rain threshold of 0.125mmh^{-1}, and given a typical rainfall rate distribution, case studies showed that POD and FAR[7] values were 72.4% and 36.8% respectively, compared to averages of case studies from UKMO mesoscale model analyses of 69.0% and 51.0% respectively (C. Jones, personal communication, 1996). The study summarized that many improvements could be made (as well as training a network using Meteosat imagery), including the replacement of the rainfall rate at the sample centre with the average rainfall rate in the sample, and non-linear neural network classification (with MLPs for example).

Using the experience gained from these pilot studies, this report considers a number of products that could be developed, with the primary aim of improving NWP performance.

A detailed account of pattern recognition techniques is perhaps inappropriate in this report, (the review published in Meteorological Applications should be suitable for further reading), but a short description of advantages over other methods is given in appendix A, with particular reference to the estimation of cloud cover amount. A brief description of the way in which MLP neural networks are trained and used is provided in appendix B; further detail can be obtained from the textbook by Beale & Jackson (1992).

2 The Products

A number of possible pattern recognition products that might be of benefit to NWP models at the UKMO are presented in table 1. The products fall into two broad categories: those that may have a potential impact in NWP models by improving existing observations (within MOPS[8] for example), and those that may provide completely new observation types. An indication of the possible area of use by UKMO divisions other than NWP is given alongside each product.

[5]National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer
[6]Global Precipitation Climatology Project Algorithm Intercomparison Project no.2
[7]Probability of Detection, False Alarm Ratio
[8]Moisture Observation Preprocessing System (see Macpherson *et al.*, 1996)

	PRODUCTS	AREA OF USE
	Improved NWP Observations	
1	Cloud cover	NWP, OS
2	Cloud type	NWP, OS
3	Synthetic relative humidity profiles	NWP, OS
4	Precipitation rate	NWP, OS
5	Radar anaprop removal	NWP, OS
6	Satellite winds	NWP
7	Sea-ice extent	NWP, OA, CR
	New NWP Observations	
8	Cloud system type (forecast error)	NWP, OS, CR
9	Surface pressure (objective PAOBs)	NWP
10	Potential vorticity	NWP, JCMM

Table 1: Ten possible pattern recognition products derived from satellite imagery, together with likely UKMO customers.

Each of these products is described in this section in more detail, including some background, required resources, likely customers and potential benefit. NWP relevant products are considered in relation to the current versions of the UKMO Unified Model (Cullen, 1993), namely the global, limited area and mesoscale forms.

The first two products considered are examples of how pattern recognition might be used to the direct benefit of the MOPS scheme in either the mesoscale model or the LAM[9]. In performing a three-dimensional analysis of cloud amount in the mesoscale model, MOPS blends a number of observations from different sources, namely the UK radar network, surface reports and Meteosat infrared satellite imagery. These result in a precipitation rate analysis, a total cloud cover analysis, cloud-top and cloud-base height analyses, and finally, a multilevel cloud analysis. Apart from radiosonde observations, MOPS is the principal method of providing observations of moisture in the atmosphere, and is therefore a convenient framework in which to assimilate satellite imagery pattern recognition products relating to moisture fields.

2.1 Cloud cover

The current method of estimating cloud cover in the mesoscale model version of MOPS is to use Meteosat infrared imagery which has been temperature-calibrated, from which pixels are assumed to be cloud filled if they are colder than the mesoscale model surface temperature by some specified threshold (Macpherson *et al.*, 1996). The number of cloudy satellite pixels (with a resolution of 7km) is counted within a mesoscale model grid square (17km resolution) to give an estimate of the total cloud cover.

[9]Limited Area Model

This technique is the same as the thresholding process described in further detail in appendix A, which is compared to a number of other methods of estimating cloud cover. The main objection to using the technique is found to result from the setting of threshold values, which are generally functions of many variables. This problem was highlighted by Macpherson *et al.* (1996): "*The threshold temperature difference in the satellite diagnosis step was found to require careful tuning. Too high a threshold can cause problems with relatively warm low cloud over a cold surface*". Also, "*The current threshold does not vary geographically or in time. This represents a weakness of the scheme*".

As described in appendix A, a superior method to thresholding is to use pattern classification with a number of input features, such as spectral and textural features, solar elevation, position in the diurnal cycle, surface elevation and so forth. This enables geographical and temporal information to be included, and also allows for a solution to the problem of detecting warm cloud over a cold surface.

It is interesting to note for example, the results of a pattern recognition technique that was used on Arctic AVHRR imagery (Ebert, 1987). The technique classified the imagery into one of 18 classes, including three low cloud types, each occurring over snow or ice: stratus, stratocumulus and cumulus (the same cloud types occurring over land or water were separate classes). The classifier used five spectral and three textural features (obtained by a feature selection algorithm based on divergence), from channels 1,2,3 and 4. The classification accuracies were obtained from a set of images taken on 1st July 1984 (independent from the training data), by comparing the computed class with a classification determined by three separate analysts. The accuracies were 80.0% for stratus over snow or ice, and 87.7% for stratocumulus over snow or ice (unfortunately, no independent results were available for cumulus over snow or ice). The overall classification accuracy for all 18 classes from these independent, manually labelled samples was 83.9%.

How would a pattern recognition technique be used in estimating the total cloud cover for MOPS? There are three possible approaches:

1. A two class cloud mask would be developed, primarily using Meteosat infrared observations to calculate features (but visible, geographical and temporal information could be included), and mesoscale model cloud field analyses to train a neural network classifier (note that model fields must be quality controlled, to avoid training the classifier on model errors). The analyses would ensure training the network with our best current knowledge of cloud cover, in a purely objective way.

   The cloud mask would be used to classify individual pixels, as with the thresholding technique. (The network would classify a pixel with a number of features taken from its immediate neighbourhood). Cloudy pixels would then be counted within a model grid square to estimate the total cloud cover, as in MOPS.

2. A neural network cloud type classifier would be trained using manually labelled Meteosat image samples (see section 2.2). A number of different cloud classes would be used (of the order of 10 or 20), each with an associated total cloud cover, typical of that class.

   In this case, features would be calculated for larger areas (of a *similar* size to the mesoscale model grid squares, but not the same so as to avoid locking the feature calculation to model grid square areas), using the satellite pixel grey-levels present, as well as temporal and geographical information. The network would assign a cloud type to each area, and each cloud type would have an associated total cloud cover (via a look-up table). For example, stratus might have a total cloud cover of 0.9, whilst several classes of cumulus might exist with cloud covers ranging between 0.1 and 0.6. The areal cloud cover estimates would then be interpolated to the appropriate model grid squares.

3. The logical extrapolation to the first two methods is to train a neural network directly from mesoscale model analyses of cloud cover. Thus a number of features would be calculated from the imagery within the local neighbourhood of a model grid square. The total cloud cover from model analyses would then be used to train a neural network classifier, which could have one output with a continuous range of cloud cover values (0.0 to 1.0), or a number of outputs, each representing a small range of cloud cover values (0.0 to 0.1, 0.1 to 0.2, ...). An advantage of the latter method is that quality control can be included in terms of the likelihood of a sample belonging to a particular class (see the discussion at the end of appendix B).

   All the required features would be calculated within the local neighbourhood of the model grid square to assign the total cloud cover. The classification would simply be a direct estimate of the total cloud cover in the model grid square, with the addition of an uncertainty. (Note however that it would be possible to obtain a measure of uncertainty in all three proposals described here).

Meteosat imagery collocated with a large number of training samples (from model analyses or manual labelling) would be required, ideally spanning a whole year, and obtained throughout periods of day and night[10]. Such data exist in SIAG's image archive, and previous experience with GANDOLF has shown that around two weeks are required to extract 365 image pairs taken during the course of a year (the time mainly occupied in reading tapes). A sufficiently large number of training samples need to be obtained to provide the network with good generalization (as a rule-of-thumb, Tovinkere *et al.* (1993) recommended using $10N_c$ to $20N_c$ samples per class, where N_c is the total number of classes).

Methods 1 and 3 would both use mesoscale model analyses as training data for the

[10]Various methods can be used in a pattern classification system to account for lack of data (in the visible channel at night): a switch between two separately trained classifiers, one trained with visible and infrared imagery, the other with infrared alone, or training such that all image data are used with visible features set to unique values during periods of dark.

satellite imagery. In method 2, only satellite imagery is required, but subjective labelling has to be performed, which can be time consuming. For the neural network developed for GANDOLF, 2190 samples were labelled, six per day for a total of 365 days during 1994. The labelling was undertaken by Clive Pierce of O(RS4) together with the author, in a total of four days. It is envisaged that about 4000 samples would be needed in this case, requiring 16 days of staff time in total. However, the time required to extract samples from suitable model fields for methods 1 and 3 is probably not dissimilar.

The features chosen to train the network can be selected by experience, or by using an objective feature selection technique. The FSET described in the introduction can be used to advantage here; an estimate of the amount of time required to perform an objective feature selection for such a network is one month. Finally, training and testing of the network must be undertaken. As a rough guide, a period of two months is suggested as a minimum. The total amount of staff time therefore comes to a minimum of around four to five months, for any of the three methods suggested.

A final note should be made concerning alternative reasons for obtaining very high quality cloud detection, namely that cloud clearing is of vital importance in determining a variety of surface parameters relevant to NWP, such as sea and land surface temperature, surface albedo, vegetation index and sea-ice extent.

2.2 Cloud type

Cloud type determination was the original reason for introducing pattern recognition techniques into the analysis of satellite imagery in the 1970s, in order to produce objective nephanalyses. Cloud type also plays an important rôle in the segmentation of satellite images into regions associated with different cloud systems (Pankiewicz, 1995); further discussion of this aspect is left for section 2.8.

However, cloud type is also needed as a constraint for the cloud cover estimates provided to MOPS (Macpherson *et al.*, 1996), specifically to assign an upper limit of 4 oktas if the cloud is cumulus or cumulonimbus, and to assign a minimum cloud thickness according to cloud type. Cloud type observations in MOPS are obtained from a limited number of surface observations, whereas automated satellite nephanalyses would enable coverage over the whole mesoscale model field area. Cloud type determination could be of use in providing the second cloud cover estimator described in the previous section.

The advantages of a pattern recognition approach result from the fact that many different features can be incorporated besides visible and infrared spectral data, such as textural, temporal and geographical information. Since no cloud type fields exist within current NWP models however, image samples would have to be labelled either with collocated ground-based synoptic reports or manually. To train a neural network

pattern classifier, features would need to be calculated from the image samples over a sufficiently large area to provide an accurate classification of the general cloud field type. As mentioned in the previous section, the areas would be similar in size to mesoscale model grid squares, but sufficiently different to avoid locking the feature calculation to the grid squares.

A number of spectral and textural features have in the past provided very accurate cloud type discriminators. For example, Lee *et al* 's (1990) MLP classifier was able to discriminate cumulus, stratocumulus and cirrus to 93% accuracy; Welch *et al.*'s (1992) MLP classifier was able to discriminate polar AVHRR scenes into water, solid sea-ice, broken sea-ice, snow-covered mountains, snow-free land, thin stratus or fog over ice, cirrus over ice, stratus over water, cumulus over water and multilayer cloudiness to 92% accuracy; and finally Bankert's (1994) probabilistic neural network classifier could discriminate AVHRR scenes into low cloud, altostratus, high cloud, "precipitation-type" cloud and a clear class with 91% accuracy.

The building, training and testing of such a cloud type classifier would need the following steps undertaken: a decision on the choice of cloud type classes to use; the gathering of Meteosat imagery either collocated with synoptic ground-based observations, or manually labelled; feature selection, and finally training and testing of the network on independent data (either synoptic reports of cloud type not used in training, or a set of manually labelled training samples not used in training). The timescale required would be of a similar amount to that of the total cloud cover product described earlier: a suggested minimum of four to five months.

2.3 Synthetic relative humidity profiles

In the previous two sections, pattern recognition methods have been suggested to improve the extraction of cloud data from satellite imagery, including total cloud cover and cloud type, so that they can be used within MOPS to arrive at a more accurate three-dimensional cloud analysis. At present, the diagnosed cloud fraction from MOPS is assimilated into the model by converting it into a set of relative humidity profiles, which are treated as pseudo radiosondes with a single grid point of influence. The main point is that the model is nudged toward the corresponding relative humidity from a parametrization between cloud fraction and relative humidity (Macpherson *et al.*, 1996).

One of the possible drawbacks of this method is that the parametrization only considers layered cloud (described earlier as dynamic cloud). As an alternative to this approach, it might be possible to relate a given cloud profile within a model grid square to a particular relative humidity profile, typically associated with such cloud. Thus a synthetic relative humidity profile is assigned to a model grid square, based on a previous statistical analysis of cloud in satellite imagery and collocated radiosonde relative humidity profiles. Instead of the cloud analysis governing the relative hu-

midity profile via one parametrization, it would result in a more realistic relative humidity profile typically found for that case. This type of scheme has been used with some success in an Australian mesoscale model (Mills & Davidson, 1987), and in the Canadian regional model (Sarrazin *et al.*, 1994), the latter using the method of Garand (1993).

Although Macpherson *et al.* (1996) noted the problem by saying that *"The UKMO model's relationship (between cloud fraction and relative humidity) is undoubtedly idealized"*, the point was also made that *"It would not be productive to assimilate cloud data into the model in a manner inconsistent with a relationship obeyed by the model cloud scheme ... To obtain a model analysis that fits the cloud observations, the assimilation scheme must respect the model parametrization as a "strong constraint"."* However, if synthetic relative humidity profiles were assimilated directly into the model, then the strong constraint would surely be bypassed, and the assimilation process would not be concerned with the parametrization between cloud fraction and relative humidity: this information would be implicit in the statistically analysed relationships between cloud in the satellite imagery used for training, and the relative humidity data from the collocated radiosonde profiles.

Recently, an attempt was made to assimilate cloud-top temperature information into the LAM by adjusting the humidity field (Richards & Whyte, 1995). Some case studies were performed for model runs during May 1994. Unfortunately, the results were inconsistent: the generation of a more realistic cloud amount in the model was found to cause unrealistic precipitation rates, particularly for low cloud, and if the assimilated cloud was inconsistent with the model dynamics, the model could have the effect of removing it. This meant that the best form of the scheme was found to be when negative relative humidity increments only were made, but then the statistics generally showed relatively little overall impact but the occasional dramatic improvement. However, this attempt relied on the use of the parametrization scheme between cloud fraction and relative humidity to nudge the model toward the "inferred" relative humidity. If cloud and humidity different to that idealized in the parametrization scheme were actually observed, then perhaps this might account for the inconsistent results observed? Richards & Whyte (1995) in fact recommended the inclusion of some form of real-time cloud detection and classification, perhaps more in keeping with the idea of using synthetic relative humidity profiles.

In previous work by Mills & Davidson (1987), 19 different dew-point depression profiles were identified, according to the layer of maximum cloudiness (five layers were considered), the cloud fraction and the cloud type (whether or not the cloud was stratiform or cumuliform). Cloud type was determined as a function of the standard deviation of cloud-top temperature within the area, with a value greater than 3.5°C indicating cumuliform cloud. If the layer of maximum cloud was below 850hPa, the cloud amount was less than 20%, and the cloud type was not cumuliform, than the area was deemed to be "clear". In some cases, the method was found to impact positively on the Australian mesoscale model (and subjectively, it appeared to reflect observed cloud features better than analyses using conventional data). They recom-

mended improving the system by also defining areas of deep convection, and adding more synthetic relative humidity profiles relating to "rarer" categories.

More recently, Garand (1993) has been able to retrieve dew-point depression profiles from Meteosat imagery at six standard levels: 1000, 850, 700, 500, 400 and 300hPa. This was achieved by collocating radiosonde profiles with Meteosat imagery over Europe from data taken during three months in 1988, using over 2000 radiosonde profiles to train a pattern recognition system. In this case, a clustering technique (the minimum squared error procedure) was used to identify distinct classes or profiles in a feature space consisting of cloud-top pressure, cloud fraction and cloud albedo. Interestingly, it was found that only seven distinct profiles were needed to contain the essential information on the cloud type of interest if infrared imagery were used alone, or nine if visible and infrared were used together. (The water vapour channel proved very useful in improving dew-point depressions at higher levels and the visible channel improved inferences of low-level humidity in classes associated with precipitation.) Expressed in terms of relative humidity, results on independent test data showed standard deviations varying from 13% to 20%, depending on level. In conclusion, it was however suggested that improvements could be made by including information from other channels (or sources), or by setting constraints aimed at improving the vertical structure of the retrievals.

If direct assimilation of relative humidity was possible in the UKMO Unified Model (perhaps using variational techniques), then a pattern recognition classifier trained to recognize one of a number of synthetic relative humidity profiles might provide an important new set of observations.

The first task would be to identify a number of distinct relative humidity profiles often seen in radiosonde data, although several tens of such profiles would not be difficult for a typical MLP network to handle. Collocation would have to be performed with Meteosat imagery, so that features could be calculated in the imagery and trained on the distinct radiosonde profiles. Thus an MLP network might be built with 20 output nodes, each representing a separate radiosonde profile. Feature selection would need to be performed in order to find the best set of features capable of discriminating each of the distinct profiles, but this could be achieved using the FSET described in the introduction.

Once the winning features had been obtained, an MLP network could be trained and tested, such that during classification, the output node with the largest value would indicate the associated relative humidity profile to use. An estimate of the uncertainty in the classification could be provided in the way described at the end of appendix B. Since this type of classifier has not been tried in any of the pilot studies, some time would be required to become acquainted with the radiosonde data, and to find out what constitutes a significant difference between profiles, so as to select the profiles and hence features that would distinguish important cases.

Such a scheme would have many similarities to precipitation rate classification dis-

cussed in the next section, and an estimate of the time required to build a prototype synthetic relative humidity classifier would be nine to twelve months.

2.4 Precipitation rate

One of the pilot studies undertaken in 1996 was to assess the possibility of building a pattern classifier, trained on ground-truth radar data, to estimate precipitation rates from satellite imagery. Past work has shown that estimation of precipitation rates from such data are notoriously difficult (WMO, 1994), with poor results often being obtained in the mid-latitudes (Jackson *et al.*, 1995).

The primary advantage of using pattern recognition techniques as an alternative to previous methods is the same as those described in appendix A: the inclusion of a number of dissimilar input features with which to train the classifier. For example, Zhang & Scofield (1994) presented results of a neural network trained to estimate heavy convective rainfall rates from satellite imagery using the following features: cloud-top temperature, a cloud growth factor, a rain burst factor, an overshooting top factor, a merger factor, a saturated environment factor, moisture correction and speed of the storm. Their average test error from a total of seven cases was 3.9%, compared to an average error from the "meteorologist-computed" Interactive Flash Flood Analyzer system of NOAA/NESDIS[11] of 30.4%.

In another recent study, Uddstrom & Gray (1996) considered the use of discriminant functions to estimate rainfall rates from different cloud types, observed in AVHRR imagery and verified with New Zealand radar data taken over the course of 1994. Their idea was to initially classify imagery into eight cloud classes, and to use separate classifiers for each of the classes likely to produce rainfall (altocumulus, cumulus, cumulonimbus, cirrus over cloud and nimbostratus). Results indicated that a rain/no-rain classifier was accurate to 60%, 64%, 85% and 93% respectively for cirrus over cloud, altocumulus, nimbostratus and cumulus (no results were presented for cumulonimbus owing to sample size limitations). An analysis was then undertaken of how different features were significant in determining rain rate. The most useful features were found to be textural in the case of altocumulus and nimbostratus, whereas cumulus was the *only* class to show a significant relationship between cloud-top temperature and rain rate.

An important innovation in the UKMO pilot study was to concentrate on feature selection before pattern classification; in other words, to determine the set of features calculated from satellite imagery which maximizes a probabilistic distance measure between all the classes under consideration.

In the pilot study, 27 sets of four and five channel AVHRR imagery were taken from

[11] National Environmental Satellite Data and Information Service

the GPCP AIP-2 study, which had been collocated with radar ground-truth data from the UK Frontiers and European COST-73 radar networks (WMO, 1994). A total of 105 different features were calculated (21 per channel) for a number of samples taken from the satellite imagery, each sample being labelled according to one of four rain rate classes, determined from the collocated radar data corresponding to the centre of the sample. The four rain rate classes were separated by boundaries of 0.125mmh^{-1}, 1.00mmh^{-1}, and 8.00mmh^{-1}. Probabilistic distance measures were calculated for all combinations of feature vectors up to and including a feature vector length of nine dimensions, beyond which it was found that probabilistic distance measures decreased, indicating that there was more "noise" associated with the addition of extra features than there was discriminating information.

This exercise resulted in a best probabilistic distance measure being obtained when the following eight features were used: the minimum, maximum and spread in channel 5, the mean angular second moment, mean and maximum contrast in channel 3 and the mean angular second moment and mean contrast in channel 5 (the latter five features all being textural). By testing a classifier trained with such features, the *average* POD and FAR values for nine of the 27 AIP-2 cases (complete image sets) were found to be $71.6\pm9.6\%$ and $4.0\pm2.0\%$ respectively, a significantly better result than previous AVHRR methods (Jackson *et al.*, 1995). Comparison was also made with the average of a number of case studies taken from the mesoscale model (C. Jones, personal communication, 1996), providing the scores shown in table 2.

	PR	T+0	T+6
POD	72.4	69.0	63.0
FAR	36.8	51.0	56.0
HKS	63.7	51.0	46.0
TS	50.9	40.0	35.0

Table 2: Comparison of scores obtained from the UKMO pilot study to estimate precipitation rates from satellite imagery using pattern recognition, and mesoscale model scores. POD = Probability Of Detection, FAR = False Alarm Ratio, HKS = Hanssen Kuiper Score, TS = Threat Score (see Jones & Macpherson, 1996). Scores are based on the percentage of pixels in the mesoscale model area, dependent on the rainfall distribution. The values quoted in the text are based on the percentage of pixels in each of the four rainfall classes. Both used a rain/no-rain threshold of 0.125mmh^{-1}.

In summary, the study showed that a useful precipitation rate classifier could be produced, relatively cheaply, and specifically tuned for use in a variety of UKMO applications. A number of systems at the Met. Office might be improved with such estimates of precipitation rate, namely:

1. Assimilation of precipitation observations in NWP. It is possible to assimilate precipitation rates at a range of scales via latent heat nudging (LHN) schemes, as has been recently shown in the mesoscale model (Jones & Macpherson, 1996).

2. Cloud analysis in NWP. The MOPS scheme uses estimates of rainfall rate from the radar network to make inferences about total cloud cover and to make adjustments to the multilevel cloud analysis. This is done by modifying the cloud profile to ensure there are two cloud free layers below the -15°C level in the absence of rain, or presence of light rain, for example (Macpherson *et al.*, 1996). These adjustments are currently only available within the area covered by the UK radar network; measurements based on satellite imagery would increase the coverage within the system.

3. Nowcasting cloud and precipitation. Nowcasting systems such as Nimrod (Golding, 1995) are clear candidates for improved observations of precipitation rate, especially as they would be able to provide enhanced coverage of the UK and surrounding waters over and above the current UK radar network. A further use of satellite-based precipitation rate would be as a means to improve the reliability of anomalous propagation echo deletion in Nimrod radar processing. This last application is discussed in the next section, but still under the broad heading of improved NWP observations, since the latent heat nudging scheme referred to earlier has recently started using precipitation data supplied by Nimrod.

Latent heat nudging is a method which was implemented in the mesoscale model in April 1996 (Jones & Macpherson, 1996), as a way of assimilating precipitation rate data. Since precipitation rate is not a prognostic variable in the model, it cannot be assimilated directly, but can be used to infer information about the model's temperature and moisture fields, in a way that nudges them so that the diagnosed rain rate fits the observations more closely. In practice, this is performed by adjusting the model latent heat profile to reflect the heating implied by the observed precipitation, effectively resulting in an enhancement (or suppression) of the diabatic heating term in the model's thermodynamic equation.

Positional accuracy of the precipitation was found to be more important in impacting the model than the rate. Since June 1996, the main source of observational data used in the LHN scheme has been provided by the UK radar network via Nimrod, and is known to be of low quantitative accuracy at ranges of over 100km from the nearest radar (Jones & Macpherson, 1995). However, Jones & Macpherson pointed out that "*Imposing this 100km range on the LHN scheme would exclude those areas on the edge of the radar range where systems are approaching the UK, systems we would hope to analyse better. The "maximum range" gives much better coverage, and it seems worthwhile, if not essential, to extract useful information from the data over this larger area, albeit recognizing the poorer quality nearer the perimeter.*" The use of satellite-derived precipitation rates would therefore appear to provide considerable benefit, either by directly providing a new observational data type to the global

model, or via Nimrod for the mesoscale model. A further advantage of using satellite derived precipitation rates (obtained with pattern recognition), is the inclusion of an uncertainty based on the precipitation rate being assigned correctly, compared to the data that the network was originally trained with, as a form of quality control.

Some of the design principles used in the LHN scheme are of relevance to precipitation estimation based on satellite imagery: independence of model resolution and domain (satellite observations could be used throughout the global model), applicability to any source of rain rate data, transparency to changes in model parametrization and portability to a 3DVar environment. With regard to its use in variational assimilation, Jones & Macpherson (1996) recommend the retention of the latent heat nudging approach in 3DVar, and point out that preliminary studies have been made using 4DVar (Zupanski & Mesinger, 1995), which demonstrate the potential of LHN in the longer term.

Although a useful precipitation estimator resulted from the UKMO pilot study, there are a number of significant improvements that can be made. The classifier was designed for use on AVHRR data, whereas for the LHN scheme in the mesoscale model, Meteosat would be the preferable data source, primarily because of the better temporal resolution (30 minutes compared to 6 hours). However, both improved temporal and spatial resolution are expected following launch of MSG[12] in 2000, with a 15 minute frequency and 3km infrared sub-satellite resolution (the high-resolution visible channel will provide 1km resolution at the sub-satellite point). To train a pattern classifier sufficiently on Meteosat data, the gathering of perhaps a year's worth of collocated satellite and radar imagery would be necessary (from one randomly chosen observation set per day), and would probably need a minimum of two months effort.

A very important part of building an accurate precipitation rate classifier is to choose the features which best discriminate the classes of interest. In the pilot study, 21 features calculated in five AVHRR channels were used, but did not include channel differences or features such as those described in the work of Zhang and Scofield (1994). A significant part of the development of a good classifier would require some three months to perform a feature selection using all the candidates of interest. This selection would be performed using FSET, which requires no further development, except to set it up for use on Meteosat imagery (perhaps one week to modify and test the software). Much of the justification for using a precipitation classifier designed in this way stems from this stage, where the ability of *any* classifier to discriminate the classes of interest is explored. Using a two stage classifier, suggested by the work of Uddstrom & Gray (1996), may produce even more accurate estimates, dependent on cloud type. This would obviously need the extra work described in section 2.2, occupying a minimum of four months.

Before incorporation into the LHN scheme, or as a new observational type in Nim-

[12]Meteosat Second Generation

rod, the actual classifier must be developed, trained and tested using the winning features. An MLP network is the preferred option, because of ease of use, the fact that a general MLP classifier has already been produced, and the possibility to provide uncertainty estimates, as well as speed in operational use and high classification accuracy compared to other techniques. The work would involve identification of the best network architecture, training using samples taken from the collocated (and quality controlled) ground-truth radar data, and testing using independent samples from the same data set. Approximately two months is estimated as a guide to the amount of staff time required. The total estimate for the development of a Meteosat based, accurate precipitation rate classifier is therefore around seven months (or eleven months with the addition of a cloud classifier).

One point should be noted about the potential use of the precipitation rate classifier in Nimrod. Although using features calculated from a local neighbourhood of Meteosat pixels, the classifier would be trained on individual 5km resolution radar data, so that the resulting precipitation rate classifier could produce a field of 5km resolution. Alternatively, resolution could be decreased to the 15km currently used in Nimrod on smoothed 5km data, for interpolation to the 17km mesoscale grid.

Finally, it should be noted that there is scope to design a pattern recognition system encompassing microwave as well as conventional visible and infrared imagery. As was shown in the AIP-2 report (WMO, 1994), methods of retrieving precipitation rate from SSM/I also suffers drawbacks in the mid-latitudes. It might be possible however, that a pattern classifier combining visible, infrared, and microwave imagery could produce a significantly better method for detecting precipitation than current methods show. Such data observed from the same platform would be an ideal way forward, and this might be feasible using the AMSU-B instrument due for launch on NOAA-K in 1997. Such research may be well worth pursuing at the UKMO, but would require a timescale of the order of a year to combine the data, do the feature selection and train a classifier.

2.5 Radar anaprop removal

As well as using satellite based precipitation rates in NWP models, or directly as a new observation type in a nowcasting system, satellite-based precipitation rate estimates might be used to quality control existing UK radar observations. For example, it has been pointed out that there is a need to improve the reliability of deleting two unwanted signals in the radar processing performed by Nimrod, namely anomalous propagation (anaprop) and clutter (M. Kitchen, personal communication, 1996). Meteosat infrared imagery is currently used to derive a probability of precipitation based on an infrared thresholding technique, which provides a POD of around 60% for cold frontal precipitation, worsening for others.

Some of the requirements for such a product include resolutions of 5km and 30 min-

utes, 95% overall availability, each product no later than 5 minutes after the satellite observation, a FAR of less than 5% within cloudy regions and a POD of more than 98% for precipitation rates greater than 1.00mmh^{-1}, also within cloudy regions.

The spatial resolution of Meteosat imagery at UK latitudes is approximately 3km (east-west) in the visible and 6km (east-west) in the infrared. (Resolutions of 5km (infrared) should become available after launch of MSG in 2000). Experience from running the GANDOLF classifier semi-operationally has shown that a classified image can be made available within 30s of receipt of both images, using a typical UKMO HP unix workstation, nearly 100% of the time (dependent on Autosat-3). The pilot study precipitation rate estimator described in section 2.4 used a rain/no-rain threshold of 0.125mmh^{-1}, and provided a one sigma FAR upper limit of 6.0% and POD lower limit of 62.0% for the nine cases studied in spring 1991. The FAR is almost within the requested tolerance, and it is likely that with a higher threshold of 1.00mmh^{-1} and improved feature selection using a larger range of candidate features, it would prove easier to obtain a higher POD than the average value of 71.6%, which was obtained within a number of different synoptic conditions.

The pilot study classifier could not simply be transposed for use in anaprop deletion, since it was trained using AVHRR data for a small number of case studies in spring 1991. Data gathering would need to take place in order to provide a sufficient number of collocated Meteosat and radar image sets to train a pattern classifier. However, since the classifier would provide a probability of precipitation rather than individual precipitation rates, the number of samples per class rule-of-thumb (Tovinkere, *et al.*, 1993) shows that 20 to 40 samples per class would be needed, although around 200 to 400 per class would be better, to capture different synoptic situations. This is because only two neural network outputs would be required, representing precipitating and non-precipitating classes, with a threshold of 1.00mmh^{-1}.

Each output would have an associated measure of uncertainty, and as a first step, the actual output value of the precipitating class could be used to provide the probability of precipitation field required for Nimrod anaprop processing. The cheapest method would be to choose about 10 or 20 collocated image sets spread throughout a year, giving different synoptic situations. Around 20 samples would then be chosen for each case to train the network. The amount of effort required in obtaining such data might be of the order of one month.

Although the same features that showed the best class separation in the pilot study could be used for a probability of precipitation field, it would be preferable to re-do the feature selection on Meteosat imagery for the new set of cases, as they would have been taken from throughout a whole year, and a significantly different threshold would be considered. The work required to use the FSET to look for the best feature separation in this situation would probably take around one and a half months.

Time would also be required to train and test the MLP network and decide on its architecture. Around one and a half months effort is seen as the minimum time

required in this case. So the total amount of staff time required to build, train and test a classifier for use in the deletion of radar anaprop would be in the region of four months. This work would however provide a useful base from which to go on and build a classifier for discriminating several classes of precipitation rate.

2.6 Satellite winds

Winds derived from satellite data are obtained at a number of different atmospheric levels by various national meteorological services, and the production of winds of particular relevance to NWP at the UKMO has recently been reviewed by Butterworth (1997).

One of the ways in which pattern recognition techniques may increase the accuracy of satellite winds is to improve the tracking of clouds throughout successive images. Currently, Meteosat derived winds (Schmetz *et al.*, 1993) use a maximum cross-correlation technique over three successive infrared images to determine a displacement vector. However, work by Côté & Tatnall (1995) has shown that a Hopfield neural network can result in more quickly processed and more accurate wind vectors.

Hopfield networks were originally designed as associative memory devices (Beale & Jackson (1992) give a good introduction). In the work of Côté & Tatnall, a number of sharp cloud edges represented by parabolic functions were identified, each of which was assigned to a node in the Hopfield network, arranged in rows (for the first image) and columns (for the second image). The network was then used to minimize an energy function which matched the edges previously identified in the images.

It may be of interest to produce a faster, more accurate satellite wind tracking product within the UKMO, based on these ideas, although it is also recognized that improved wind height assignment is likely to lead to a larger model impact in the short-term. To improve tracking using the pattern recognition technique, some time would be spent building a prototype network and testing the method on half-hourly Meteosat imagery using data from the Autosat-3 system. It is expected that only around four months would be needed to build a prototype. Incorporation of in-house satellite winds using this approach, with the appropriate quality checking would obviously require further effort.

2.7 Sea-ice extent

An observed boundary of sea-ice is used in the Unified Model, and is obtained from the U.S. Navy/NOAA Joint Ice Centre. These data are based on AVHRR and passive microwave satellite observations, and are obtained on a weekly basis. If further detail were required (either temporally or spatially), it would be possible to build a classifier

similar to that of Ebert's (1987) to identify areas of open water, unbroken sea-ice, broken sea-ice, and ice sheet besides cloud and land cover from AVHRR imagery, at around 30km resolution. In that study, spectral and textural features were used to obtain classification accuracies of the order of 95% for those surface types.

Since experience has been obtained in producing cloud classifiers similar to the type required to detect sea-ice, the main emphasis of the work would be in obtaining training data, which could be manually labelled or collocated with surface reports. Such a project would be expected to last about six months, before a prototype sea-ice classifier could be demonstrated.

2.8 Cloud system type (forecast error)

One of the original reasons for undertaking pattern recognition research in SIAG was to see if cloud systems (such as cold fronts and depressions), as opposed to cloud fields (stratus and cumulus for example), could be identified from satellite imagery.

Previous cloud system classification techniques have been directed toward the automatic interpretation of imagery for use by forecasters (although only a few examples of such work are known in the literature). The scheme of Peak & Tag (1992), was designed as part of an expert system for providing meteorological information to U.S. Navy ship-based forecasters, such as the likelihood of clear air turbulence. They used the subjective approach of asking a meteorologist to identify 172 synoptic scale objects in GOES-West satellite imagery, which were later binned into eight separate types. For each object, the amount of low and multilayer cloud was calculated, as well as the longitudinal and latitudinal extent, and the northernmost latitude. An MLP network was then trained from 63 of the objects, providing a classification accuracy of 83% on the remainder. Unfortunately, the cloud systems had to be manually segmented from the imagery by a meteorologist, but the authors pointed out that a bottom-up approach (to segment the image into small regions first, followed by larger regions) might be used to identify specific synoptic-scale systems in the image.

More recently, Zwatz-Meise *et al.* (1996) have been attempting to automate the so-called "Satreps" produced by the Central Institute of Meteorology and Geodynamics (ZAMG) in Austria. These are routine, daily analyses of satellite images, in which conceptual models are manually identified, primarily to allow forecasters to subjectively evaluate the quality of NWP analyses. The automated version involves an initial cloud classification using infrared brightness temperature and the number of local extrema as features, to identify two characteristics: whether the cloud is low, medium or high, and whether it is homogeneous, lumpy or cellular. Similar and adjacent samples are then connected into larger regions, from which attribute fields are calculated, each belonging to a specific conceptual model. At the time of writing, such attribute fields had been identified for cold fronts, warm fronts and jet streams.

Efforts to develop a scheme to identify cloud systems at the UKMO were largely based on the ideas of Peak & Tag (1992), of providing a classifier that can segment an image in a bottom-up fashion. It became apparent that different features should be calculated at different scales, depending on the pattern that a classifier was being trained to recognise. For example, spectral features should be calculated for small groups of pixels (or individually) for cloud amount, textural features should be calculated over regions of perhaps tens to hundreds of kilometres (for cloud fields), and spatial features at the synoptic scale (for cloud systems) should be calculated once large areas of similar cloud type had been identified (Pankiewicz, 1995). This led to the idea of a three-level cloud system classifier, for which features calculated at each of three scales would provide the most important information to discriminate classes found at those scales. It is interesting to see that these ideas are similar to the work of Zwatz-Meise *et al.* (1996), who use simple spectral and textural features to identify cloud type, and larger scale attribute fields to relate conceptual models to spatial structure and meteorological variation at the synoptic-scale.

However, the interest at the UKMO is to provide new observation types from pattern recognition of satellite imagery that may be used to influence NWP models, by the way in which synoptic-scale structure is contained within the model. Some recent work by Hoffman *et al.* (1995) has shown that the identification of such structures may be very well suited to reducing forecast errors (background minus analysis) within a variational analysis framework. They considered the decomposition of forecast errors in terms of three components: a displacement error (how much of the forecast error can be accounted for by moving the forecast to best fit the analysis), an amplitude error (how much of the forecast error can be accounted for by changing the amplitude of the displaced forecast, which results in a combined distortion error), plus a residual error (anything unaccounted for by distortion). Their basic premise was that forecast errors are best described by reference to large-scale meteorological objects, so that generally, small-scale errors can be attributed to the residual error. Determining the distortion errors in this way is a problem of minimization, and Hoffman *et al.* argued that in a variational framework, this methodology can accommodate any type of data which can be diagnosed from the model forecast.

Many types of satellite data have potential for correcting distortion errors, but have only had limited use in conventional data assimilation systems, for example cloudiness in synoptic-scale systems, precipitable water (the latter often showing very good agreement with frontal analyses) and water vapour. In practice, Hoffman *et al.*'s study used precipitable water and surface wind speed from SSM/I and ERS-1 backscatter, to demonstrate the detection and characterization of forecast errors in terms of position and amplitude errors (compared with ECMWF operational analyses), by means of a cross-correlation technique.

In terms of assimilating these errors into a variational scheme which might improve NWP analyses, they had to choose a distortion which minimizes an objective function. The objective function used incorporates the sum of the residual and distortion penalty functions (J_r+J_d), replacing the usual background penalty function J_b, as well

as the normal observation cost function J_o and other constraint penalty functions J_c:

$$J = J_r + J_d + J_o + J_c \tag{1}$$

Such an objective function is more complex than in the standard procedure (Lorenc, 1986), and more degrees of freedom are added to the variable X that we wish to analyse. This is compensated not by additional information in the observations, but by additional information in the statistics, contained within the constraint J_d:

$$J_d = C^T S^{-1} C \tag{2}$$

Here, C is a vector representing the distortion between X and the background field X_f, and S is the estimated covariance of the distortion variables. The residual penalty function is given by:

$$J_r = (X_d - X)^T R^{-1} (X_d - X) \tag{3}$$

where X_d is the distorted forecast, and R is the estimated covariance of the residual errors, expected to be much simpler (and have less structure) than the normal covariance matrix of the forecast errors, so that it can be approximated by a diagonal matrix. The aim is therefore to choose the analysis X_a and the distortion C which minimize the objective function J (so that X becomes the analysis X_a when J has been minimized).

Clearly, there will be a great deal of effort required in identifying cloud systems in satellite imagery, and estimates of distortion errors in the corresponding objects in UKMO NWP models. Large-scale meteorological objects such as fronts, cyclones, anticyclones, polar lows and waves might be the candidates, and they would need to be identified in both the satellite imagery and the model fields to be able to form distortion errors of the type described in Hoffman *et al.*'s scheme.

Methods of identifying fronts in model fields have recently been explored at the UKMO (T. Hewson, personal communication, 1996), in which several derivatives of a temperature field are obtained, together with a number of other criteria to define the position of a front; such a scheme could perhaps be adapted to this situation, although it is the cloud that is being observed in the satellite imagery rather than the temperature field. Model cloud fields (such as those regularly produced by the mesoscale model) could be used to identify the synoptic-scale cloud structures for direct comparison with the same structures in the satellite imagery. This would provide a first step to reducing forecast error relating to cloud structure, and with the potential shown for exploitation in variational assimilation schemes, these products

may well be worth investing further research and development over the next few years.

2.9 Surface pressure (objective PAOBs)

For around 25 years, PAOBs have been produced by the Australian Bureau of Meteorology, and have been disseminated to the meteorological community on the Global Telecommunications System (Seaman *et al.*, 1993). PAOB is an abbreviation for "paid observation", because when they were first used in the early 1970s, the numerical analysis programs in operation were able to fit the PAOBs closely; they were guaranteed or "paid" observations. The current assimilation systems assign observation errors, and they may be rejected by quality control procedures, so that now, they do not pay their way quite as well!

The idea of a PAOB is to extract point estimates of sea-level pressure from manually analysed charts in conjunction with satellite imagery interpretation, to identify sea-level pressure values in a regular pattern with a spacing of 1000-1500km, together with a number of additional points locating depression centres, troughs, ridges and so on. They are used by a number of global modelling centres, including ECMWF, who assign observation errors of 4hPA and still continue to find a modest positive impact in the southern hemisphere mid-latitudes (Seaman *et al.*, 1993; G. Kelly, personal communication, 1995). (They are currently not in use at the UKMO.)

Using some of the ideas described in the previous section on cloud system recognition, it would seem that depression centres, troughs and ridges could be automatically identified using a pattern recognition system. The problem would be in assigning the associated sea-level pressure, but if a similar technique were used on a pressure analysis from the global model, the corresponding pressure could be reported, but at the new location defined by the satellite observation and pattern recognition method.

The work would involve the identification of edges or centres of depressions, so that fewer observations would probably be reported than classical PAOBs. This would mean initial recognition of cloud systems, possibly via edge detection to define the system rather than cloud type, before the extraction of points in the synoptic flow. Once identified, these points would need to be compared with the same points in a model sea-level pressure field. Since PAOBs are already being produced for the southern hemisphere, greater impact might result by obtaining objective PAOBs in the mesoscale model area around the UK. If a significant departure from the pressure field was found in the observations, the model would be nudged toward the new pressure distributions in the usual way. Alternatively, the departure from the pressure field may be viewed as a forecast error, so that the distortion techniques described in the previous section would also be applied here. This would make objective PAOBs another possible candidate for variational assimilation.

2.10 Potential vorticity

Recent work at the JCMM[13] has shown that the so-called dry intrusions of air descending from the upper troposphere and lower stratosphere into the middle and lower troposphere (often seen as "dark zones" in Meteosat water vapour imagery) can be related to the rolling-up of part of a streamer of anomalously high potential vorticity (PV), which accompanies local downward-folding of the tropopause (Browning *et al.*, 1996).

The dark zone in the water vapour imagery appears to correspond to a minimum of about 6km in the height of the PV = 2 surface, representing the dynamical tropopause (Hoskins *et al.*, 1985). Regions where the tropopause descends to low levels therefore corresponds to an upper-level maximum of PV. It may therefore be possible to use Meteosat water vapour imagery to gain an understanding of PV in model areas of interest, especially as the imagery is measuring water vapour near the tropopause. Mansfield (1996), for example, presented several case studies showing how a mismatch between PV and water vapour imagery might be indicative of model error.

This work is very much at its early stages, and so collaboration with JCMM would be paramount to understand the exact relationships between PV and water vapour. For example, does the relationship hold true in most cases for it to be used operationally in a NWP model, and would the direct assimilation of water vapour imagery within a 4DVar scheme be the preferred option?

However, if the relationship is general and is suited to model assimilation, it might be possible to develop pattern recognition techniques which would automatically identify the water vapour "dark zones", and if trained with model PV field data, could supply a new observational data type implying PV structure in the model.

This area of research might suit a pilot study of around six months which would use collocated Meteosat water vapour imagery and global model (or LAM) PV fields to investigate the correlation and look for possible forecast impact.

3 Summary

Ten separate pattern recognition products have been described, all of which would be of relevance to NWP. Some products might be constructed on a timescale of a few months (cloud type, precipitation rate and radar anaprop removal), whereas some of the more challenging projects (such as decreasing forecast error with cloud system observations and potential vorticity mapping from water vapour imagery) must be viewed on a longer timescale.

[13]Joint Centre for Mesoscale Meteorology

In addition to use in NWP, many other diagnostic pattern recognition products might be worth considering. These could include automated nephanalyses (cloud fields), descriptions of cloud system types and conceptual models (similar to that of Zwatz-Meise *et al.*, 1996), fog estimation, recognition of cloud ship wakes and aircraft contrails to name a few. Unfortunately, it has not been possible to describe these and similar products within the scope of this report.

This overview is presented as a starting point for discussion on the best way of using these techniques to improve UKMO NWP model performance. Suggestions concerning the products described, as well as new ideas or alternative methods, will of course be welcome.

4 References

Arking A. & Childs J.D. 1985, "Retrieval of Cloud Cover Parameters from Multispectral Satellite Images", *J. Climate Appl. Meteorol.*, **24**, pp.322-333.

Bankert R.L. 1994, "Cloud Classification of AVHRR Imagery in Maritime Regions Using a Probabilistic Neural Network", *J. Appl. Meteorol.*, **33**, pp.909-918.

Batten S.M. 1996, "Precipitation Estimation Study", EDS Defence Ltd Technical Report no. C314470/DM3 commissioned for the UK Meteorological Office.

Beale R. & Jackson T. 1992, *Neural Computing: An Introduction*, (Bristol: IOP Publishing).

Browning K.A., Roberts N.M. & Sim C.S. 1996, "A Mesoscale Vortex Diagnosed from Combined Satellite and Model Data", *Meteorol. Appl.*, **3**, pp.1-4.

Butterworth P. 1997, "Investigation into Satellite Winds - I. Production", *Forecasting Research Technical Report*, (Bracknell: UK Meteorological Office) (in the press).

Coakley J.A. & Bretherton F.P. 1982, "Cloud Cover from High-Resolution Scanner Data: Detecting and Allowing for Partially Filled Fields of View", *J. Geophys. Res.*, **87(C7)**, pp.4917-4932.

Collier C.G., Hardaker P.J. & Pierce C.E. 1995, "The GANDOLF system: An automatic procedure for forecasting convection", *Proc. 27th Conf. on Radar Meteorology, Vail, Colorado (Boston: American Meteorological Society)*, pp.606-608.

Côté S. & Tatnall A.R.L. 1995, "A Neural Network Based Method for Tracking Features from Satellite Sensor Images", *Int. J. Rem. Sens.*, **16**, pp.3695-3701.

Cullen M.C. 1993, "The Unified Forecast/Climate Model", *Meteorol. Mag.*, **122**, pp.81-94.

Desbois M., Seze G. & Szejwach G. 1982, "Automatic Classification of Clouds on Meteosat Imagery: Application to High-Level Clouds", *J. Appl. Meteorol.*, **21**, pp.401-412.

Ebert E. 1987, "A Pattern Recognition Technique for Distinguishing Surface and Cloud Types in the Polar Regions", *J. Climate Appl. Meteorol.*, **26**, pp.1412-1427.

Garand L. 1993, "A Pattern Recognition Technique for Retrieving Humidity Profiles from Meteosat or GOES Imagery", *J. Appl. Meteorol.*, **32**, pp.1592-1607.

Golding B.W. 1995, "Rainfall-analysis and Short Range Prediction", *in Hydrological Uses of Weather Radar, ed. K.A. Tilford, occasional paper no.5*, (London: British Hydrological Society).

Hoffman R.N., Liu Z., Louis J.F., & Grassotti C. 1995, "Distortion Representation of Forecast Errors", *Mon. Wea. Rev.*, **123**, pp.2758-2770.

Hoskins B.J., McIntyre M.E. & Robertson A.W. 1985, "On the Use and Significance of Isentropic Potential Vorticity Maps", *Q. J. R. Meteorol. Soc.*, **111**, pp.877-946.

Jackson P., Holpin G.E., Foot J.S. & Allam R. 1995, "The Detection of Midlatitude Precipitation from AVHRR and SSMI: Results from a Comparison of Algorithms", *Proc. 1995 Meteorol. Sat. Data Users' Conf., Winchester, UK (Darmstadt: EUMETSAT)*, pp.401-408.

Jones C.D. & Macpherson B. 1995, "A Hydrology Correction Scheme for the Mesoscale Model Using Observed Precipitation Rates", *Forecasting Research Technical Report no. 151*, (Bracknell: UK Meteorological Office).

Jones C.D. & Macpherson B. 1996, "A Latent Heat Nudging Scheme for the Assimilation of Precipitation Data into the Mesoscale Model", *Forecasting Research Technical Report no. 194*, (Bracknell: UK Meteorological Office).

Kidder S.Q. & Vonder Haar T.H. 1995, *Satellite Meteorology: An Introduction*, (London: Academic Press).

Lee J., Weger R.C., Sengupta S.K. & Welch R.M. 1990, "A Neural Network Approach to Cloud Classification", *IEEE Trans. Geosci. & Rem. Sens.*, **28(5)**, pp.846-855.

Lorenc A.C. 1986, "Analysis Methods for Numerical Weather Prediction", *Q. J. R. Meteorol. Soc.*, **112**, pp.1177-1194.

MacKay D.J.C. 1992, "A Practical Bayesian Framework for Backpropagation Networks", *Neural Computation*, **4**, pp.448-472.

Macpherson B., Wright B.J., Hand W.H. & Maycock A.J. 1996, "The Impact of MOPS Moisture Data in the UK Meteorological Office Mesoscale Data Assimilation Scheme", *Mon. Wea. Rev.*, **124**, pp.1746-1766.

Mansfield D.A. 1996, "The Use of Potential Vorticity as an Operational Forecast Tool", *Meteorol. Appl.*, **3**, pp.195-210.

Mills G.A. & Davidson N.E. 1987, "Tropospheric Moisture Profiles from Digital IR Satellite Imagery: System Description and Analysis/Forecast Impact", *Aust. Met. Mag.*, **35**, pp.109-118.

Pankiewicz G.S. 1995, "Pattern Recognition Techniques for the Identification of

Cloud and Cloud Systems", *Meteorol. Appl.*, **2**, pp.257-271.

Pankiewicz G.S. 1996, "Towards Synoptic Scale Cloud Classification: Combining Spectral and Textural Features", *Proc. 1996 Meteorol. Sat. Data Users' Conf., Vienna, Austria (Darmstadt: EUMETSAT)*, pp.111-118.

Pankiewicz G.S. 1997a, "Neural Network Classification of Convective Airmasses for a Flood Forecasting System", *Int. J. Rem. Sens.*, (in the press).

Pankiewicz G.S. 1997b, "New Meteosat Pattern Recognition Products for Use in Weather Forecasting", *Adv. Space Res.*, (in the press).

Peak J.E. & Tag P.M. 1992, "Toward Automated Interpretation of Satellite Imagery for Navy Shipboard Applications", *Bull. Amer. Meteor. Soc.*, **73(7)**, pp.995-1008.

Pierce C.E., Collier C.G. & Hardaker P.J. 1995, "Forecasts of heavy convective rain for use within a flash flood warning system", *Proc. 5th British Hydrological Society Symposium, Edinburgh, UK (London: British Hydrological Society)*, pp.4.17-4.23.

Richards N.S. & Whyte K.W. 1995, "Experimental Use of AVHRR Imagery in Numerical Weather Prediction Data Assimilation", *Proc. 1995 Meteorol. Sat. Data Users' Conf., Winchester, UK (Darmstadt: EUMETSAT)*, pp.299-306.

Rumelhart D.E., Hinton G.E. & Williams R.J. 1986, "Learning Representations by Back-Propagating Errors ", *Nature*, **323**, pp.533-536.

Sarrazin R.J., Mailhot B., Bilodeau B., Brunet N., Methot A. & Pellerin G. 1994, "The Canadian Regional Forecast System: An Evaluation of the New 50km Version", *Preprints of the 10th Conf. on Numerical Weather Prediction, Portland, Oregon (Boston: American Meteorological Society)*, pp.420-422.

Schmetz J., Holmlund K., Hoffman J., Strauss B., Mason B., Gärtner V., Koch A. & Van De Berg L. 1993, "Operational Cloud-Motion Winds from Meteosat Infrared Images", *J. Appl. Meteorol.*, **32**, pp.1206-1225.

Seaman R., Steinle P., Bourke W. & Hart T. 1993, "The Impact of Manually Derived Southern Hemisphere Sea Level Pressure Data upon Forecasts from a Global Model", *Weather and Forecasting*, **8**, pp.363-368.

Tovinkere V.R., Penaloza M., Logar A., Lee J., Weger R.C., Berendes T.A. & Welch R.M. 1993, "An Intercomparison of Artificial Intelligence Approaches for Polar Scene Identification", *J. Geophys. Res.*, **98(D3)**, pp.5001-5016.

Uddstrom M.J. & Gray W.R. 1996, "Satellite cloud classification and rain-rate estimation using multispectral radiances and measures of spatial texture", *J. Appl. Meteor.*, **35**, pp.839-858.

Welch R.M., Sengupta S.K., Goroch A.K., Rabindra P., Rangaraj N. & Navar M.S. 1992, "Polar Cloud and Surface Classification Using AVHRR Imagery: An Intercomparison of Methods", *J. Appl. Meteor.*, **31**, pp.405-420.

WMO 1994, *Workshop Report on the GPCP Second Algorithm Intercomparison*

Project (AIP-2) for the World Climate Research Programme, Reading, UK (Geneva: WMO).

Zhang M. & Scofield R.A. *1994, "Artificial Neural Network Techniques for Estimating Heavy Convective Rainfall and Recognizing Cloud Mergers from Satellite Data", Int. J. Rem. Sens.,* **15**, *pp.3241-3261.*

Zupanski D. & Mesinger F. *1995, "Four-Dimensional Variational Assimilation of Precipitation Data", Mon. Wea. Rev.,* **123**, *pp.1112-1127.*

Zwatz-Meise V., Scheiber J. & Zobl Z. *1996, "Towards Automatic Meteosat Image Interpretation"", Proc. 1996 Meteorol. Sat. Data Users' Conf., Vienna, Austria (Darmstadt: EUMETSAT), pp.97-104.*

A Advantages of Pattern Recognition

In order to understand why pattern recognition techniques are potentially useful for retrieving a number of meteorological parameters, it is instructive to examine a particular example, that of cloud cover, and the various methods that have been used over the years. A summary of the different techniques is given below, based on a review from the textbook "Satellite Meteorology" by Kidder & Vonder Haar (1995).

The simplest approach to estimating cloud cover is to assign threshold values for both infrared brightness temperature and visible albedo, defining cloud to occur on one side of the threshold. However, cloud cover can be overestimated if the threshold is set near to the true clear radiance (in which the image sample is completely clear) and underestimated if set near to the true cloudy radiance (in which the sample is completely cloudy). Perhaps the main objection to thresholding is that the setting of such values, which can be functions of many variables (surface type, surface condition, recent weather, season, time of day and viewing geometry), is nearly always performed manually.

An alternative method is to relax the assumption that thresholds must be used in spectral space, for example by using a histogram technique such as dynamic clustering (Desbois *et al.*, 1982). Problems encountered here are that some cloud or surface types are too variable to form a local maximum in the histogram. This being an unsupervised method, cluster centres have to be identified either manually, by setting thresholds, or by comparing to previous classifications.

In multispectral retrieval techniques, radiance observed in a visible channel is modelled as solar radiation reflected from cloud tops and the surface, and infrared radiance is modelled as surface, cloud, and surface-through-cloud emission. Retrievals tend to be rather noisy, and error analyses indicate that cloud amount must be obtained from at least 50 pixels. They also require a knowledge of bidirectional reflectance, which is a function of size, shape and spacing of clouds.

For a small sample of pixels, the standard deviation of clear and completely cloudy areas can be low. The spatial coherence method (Coakley & Bretherton, 1982) constructs an arch of points obtained from each such sample in a 2D feature space consisting of visible standard deviation and infrared brightness temperature. The average radiances in the "feet" of the arch are used to determine cloud amount via an equation relating the total radiance to the individual cloudy and clear radiances. However, when cloud systems are too complex (near fronts and when cirrus overlies lower cloud), the method fails because the characteristic arch pattern is missing.

Radiative transfer techniques are often used to determine cloud optical depth and microphysical parameters, after cloud has been detected in the image of interest. The method of Arking & Childs (1985), however, selects one of six microphysical models as being the most representative

in a sample, based on a comparison of the calculated and observed 3.7μm radiance in the sample. Clear radiance values and tabulated functions for the chosen microphysical model are then used to determine a visible optical depth, a cloud top temperature and a cloud amount. The weakest part of this approach is the selection of the microphysical model, which not only requires a number of assumptions about drop size and phase, but is also rather subjective.

A better technique would be to construct a multidimensional feature space which has as its axes not just spectral features, but also texture, time, surface conditions and so on, depending on those features which are most significant in accurately classifying cloud amount (by means of feature selection). Classes of interest (such as clear and cloudy, or different ranges of cloud amount) can then be identified in this space using a suitable training set: these are the basic ideas behind the pattern recognition method.

This method is something like thresholding, but much more sophisticated. The "thresholding" functions need not be linear, but any shape determined by the training data, and information concerning surface conditions, time of year or day and viewing geometry can be accommodated by choosing the correct features on a large enough training set. Multispectral and spatial coherence methods automatically become a part of the pattern recognition process if those features are deemed important, but their disadvantages disappear because many other features are used also.

Supervised pattern recognition also avoids histogramming problems: it is not cluster centres that are important, rather regions in between thresholding boundaries. If a supervised training set is used, there is no need to label the classes after processing. Finally, no determination of a cloud microphysical model is required; the information is implicit between the selected features and the classes of interest. This means that pattern recognition techniques can also be used to explore relationships between image features and classes of interest (for example, the relationship between textural measures and rainfall rate).

B Multilayer Perceptron Neural Networks

As remarked in the introduction, multilayer perceptron neural networks provide fast, robust and highly accurate pattern classifiers, and are becoming more and more accepted in the remote sensing community. Conventional MLP networks consist of a number of input nodes (representing the features calculated in the image), one or two "hidden layers" of nodes and a layer of output nodes, the latter representing the output classes of interest. The nodes in each layer are connected to all nodes in the next layer, so that for a network with 5 input nodes, 10 hidden nodes and 3 output nodes, there are a total of 80 connection weights, each representing connection strengths between various nodes.

MLPs can be trained using the so-called backpropagation learning algorithm (Rumelhart *et al.* 1986), where an error function is calculated at each output node. For a given training sample belonging to a specific class, the feature values are calculated and are input to the network. Initially the connecting weight values are set randomly, and the error function is obtained from the difference between the true output values for the class of that sample and the estimated values. The error function is backpropagated through the network in order to update the interconnecting weights between each node. In this way, the error function decreases to a minimum after repeated presentation of training samples, each presentation being referred to as an epoch. For a network with two active layers j and k, the error function at the output layer k is:

$$\delta_k = \sigma a_k(1 - a_k)(o_k - a_k) \tag{4}$$

where σ measures the spread of the thresholding function, a_k is the estimated network output value for each class and o_k is the true output value for each class. For the preceding layer, the error function is:

$$\delta_j = \sigma a_j (1 - a_j) \sum_k (\delta_k w_{jk}) \tag{5}$$

The weights between layers k and j are often updated with the following simple gradient descent algorithm, the technique used to search for the minimum value of the error function:

$$w_{jk}(t+1) = w_{jk}(t) + \rho \, \delta_k a_j \tag{6}$$

ρ being the rate of adaption or the learning rate. For weights between i and j, the same equation was used with j replaced by i, k replaced by j, and so on.

Numerous parameters affect the performance of such a network, and those that can be varied include the number, type and normalisation of features calculated in the image samples, the number of hidden layers and nodes, the range of output values o_k used during training, the spread of the sigmoidal thresholding function σ, the number of training samples presented at each epoch and the rate of adaption of weights ρ.

The sigmoidal thresholding function is used to control the effect of very large or small feature values, in which the output or activity on a node at layer k is given by:

$$a_k = \frac{1}{1 + e^{-\sigma I_k}} \tag{7}$$

where I_k is the net input to the node from the previous layer j. The value of the spread σ determines the range of weights which are assigned to the network, which should ideally be kept within sensible values.

One of the specific advantages of such network classifiers is that values obtained at the output nodes can be used to obtain a measure of confidence in class assignment. If the highest output node value is small compared to the values used during training, the network may be suffering from poor generalisation (the ability to successfully classify samples that were not presented during training). Cost is often a factor in producing a good solution, but it may be that extra time or additional training data cannot be obtained. In that case, the maximum output value can be used to quality control the network estimate.

The output node values of an MLP network are not strictly representative of the probability of a sample belonging to a particular class. They are estimates which become less representative the further away the test sample is from the training samples in feature space. A modified method of training MLP neural networks uses a Bayesian framework (MacKay, 1992), so that the output values are much closer to the true probabilities of obtaining a particular class throughout feature space. Ideally, these modifications would be incorporated into all of the pattern classifiers described in this report.