Numerical **Weather Prediction**

Forecasting Research
Technical Report No. 266

Removal of spurious radar echoes with a Meteosat neural network precipitation classifier

by

G Pankiewicz, C Johnson and D Harrison

May 1999

The Met.Office

Excelling *in weather services*

Forecasting Research
Technical Report No. 266

Removal of Spurious Radar Echoes with a Meteosat Neural Network Precipitation Classifier

by

George Pankiewicz, Colin Johnson & Dawn Harrison

May, 1999

NWP Division
Room 344
Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

Removal of Spurious Radar Echoes with a Meteosat Neural Network Precipitation Classifier

Contents

1 Summary

This report describes the development of a new neural network scheme based on the use of Meteosat infrared, or visible and infrared imagery, to determine a probability of precipitation within the Nimrod domain, for use in removing spurious echoes from Nimrod radar composite images.

The neural networks were trained on samples taken from 3200 boxes of 17x17 pixels, selected from 48 sets of Meteosat and radar images, taken throughout the period July 1995 to June 1997 during day or night. Feature selection was performed on the images to look at the ability of various visible and infrared features calculated over different sized regions to discriminate rain from no rain at a threshold of $\frac{1}{32}$mmh^{-1}. The Meteosat features which were found to best discriminate rain from no rain in individual 5km radar pixels were the central value, the minimum, maximum, range and the ratio of maximum to minimum of 7x7 pixel infrared samples. When visible reflectivity data were also available, the best features were found to include the same features for the visible samples, except the central value.

Two multilayer perceptron neural network classifiers were trained using the features obtained, and were found to provide overall probabilities of detection of 66% for infrared and 69% for visible and infrared features, for a 1:2.8 wet to dry ratio of pixels used for testing. The corresponding false alarm rates were found to be 57% and 41%. These can be compared with the estimates for the Meteosat step of the current Nimrod anaprop removal scheme of 61% and 47% for cold frontal cases, but of 28% and 74% respectively for cold-air convection.

The new neural network scheme was trialled during summer and winter 1998, and although it was found to provide better diagnosis of anaprop during summer, and better diagnosis of rain during winter, it provided a slightly worse diagnosis of rain during summer and a worse diagnosis of anaprop during winter. Part of these problems may be associated with the lack of surface temperature data used in the new scheme (affecting the winter scores), the way in which the rain is deleted near strong cloud edges, but also in the way that the threshold for deleting anaprop has been determined (outside the neural network scheme) for the infrared neural network. As a result of these trials, the decision has been made not to implement the neural network scheme at this stage.

The major contender for improving anaprop deletion is to convert the radar network to Doppler radars. However, this is unlikely to happen in the next 5 to 10 years. As it is thought that nearly all of the problems with the neural network scheme noted during the Nimrod trial can be remedied, thereby resulting in a superior anaprop removal technique, it is recommended to continue with the development of the scheme, for example by including surface temperature, snow cover and surface climatological albedo fields, all of which are used in the current scheme.

2 Introduction

Spurious echoes in the form of anomalous propagation (anaprop) and clutter can provide fairly common, yet unwanted signals in weather radar data, resulting in false observations of precipitation.

The Nimrod nowcasting system (Golding, 1998), used at the Met. Office to generate analyses and forecast guidance in the time range 0-6 hours for the UK and surrounding areas, produces a variety of cloud and precipitation fields which are also used in initialising the mesoscale model. These precipitation data come largely from the network of radars situated around the British Isles. The quality of radar data therefore has an important consequence for both short-range forecast guidance and mesoscale model initialisation.

Currently, the Nimrod nowcasting system attempts to detect spurious radar echoes automatically using a scheme that includes a Probability of Precipitation (PoP) field determined from two independent sources: ground-based synoptic data and Meteosat visible and infrared imagery. The Meteosat data are analysed on a pixel-by-pixel basis to help form the PoP field by using a simple thresholding technique, described by Cheng *et al.* (1993), and based on the method of Lovejoy and Austin (1979). Essentially, cloud-free areas result in a low PoP, whilst in cloudy areas, PoP is estimated according to the climatological occurrence obtained by Cheng *et al.* The scheme is often referred to as the Nimrod anaprop removal scheme, although the aim is to remove all spurious echoes; in this report, references to anaprop include all types of spurious echoes.

This Meteosat PoP is then combined with the ground-based synoptic reports and forecast PoP using Bayes' theorem (Pamment and Conway, 1998), and results in a value of a parameter known as alpha, which can range from 0 (definitely no rain) through 1 (the climatological probability of rain) to large values (definitely rain). A threshold is set within the field of alphas to remove radar echoes where there is a sufficiently low probability of rain. A detailed description of the Nimrod anaprop removal scheme has been given by Johnson (1998).

Although the current Meteosat thresholding scheme provides a Probability of Detection of about 60% for cold frontal precipitation where rain rates are greater than $\frac{1}{32}$mmh^{-1}, it is about 50% for warm fronts, and only 30% for cold-air convection.

In 1996, a study was undertaken by the Satellite Image Applications Group to estimate precipitation rate in 4 classes from NOAA AVHRR data using a neural network classifier, which provided a new capability of incorporating spectral and textural image characteristics. The results were encouraging, with an average PoD of 72% at a threshold of $\frac{1}{8}$mmh^{-1}, taken in various synoptic conditions, compared to similar mesoscale model PoD values of 69% at T+0 and 63% at T+6 (Chris Jones, personal communication). False Alarm Rates were 37% from the neural network, compared

to 51% from the mesoscale model at T+0 and 56% at T+6.

The combination of different discriminatory inputs such as infrared brightness temperature and visible reflectivity texture meant that local neighbourhood information around the pixel of interest could be used to improve the estimate of precipitation, over and above that of a threshold technique (for example in cases of cold-air convection, by recognizing convective cells). The neural network approach used in this case had further advantages in that its output values provided Bayesian estimates of the PoP directly, and that it could operate at high speed, because of the statistical nature of the trained neural network.

Following these initial results, the potential was seen for developing a neural network method of diagnosing PoP from satellite imagery for Nimrod. A set of requirements was established by the Observational Products Group, listed below in table 1.

CHARACTERISTIC	SPECIFICATION
Product	Nimrod field of PoP in each pixel
Domain	Nimrod domain
Resolution	5km, 30 minutes
Timeliness	No later than HH+5, HH+35
Availability	95% overall
FAR	$\leq$0.50 for cloudy pixels in all synoptic conditions
PoD	$\geq$0.98 for rain rates $>$ 1mmh^{-1}

Table 1: *Nimrod requirement for the diagnosis of Probability of Precipitation from satellite data. (FAR = False Alarm Rate, PoD = Probability of Detection.)*

Unfortunately, the use of high resolution polar orbiter imagery from AVHRR is limited, due to its poor temporal sampling. Meteosat is however capable of providing imagery within the Nimrod domain at the required 5km and 30 minute resolution.

A neural network PoP classifier, using spectral and textural information from Meteosat, was therefore considered a suitable candidate to improve upon the current Meteosat thresholding scheme. A neural network using infrared imagery alone could be used, as such data would be available 24 hours per day. However, in practice, the inclusion of textural features from visible reflectivity data were expected to improve the network performance during daytime, and in fact, it proved to be a relatively simple matter to consider the merits of both classifiers, once the necessary training data had been restored from archive.

This report describes the production of the infrared as well as the visible and infrared PoP classifiers: how the training and validation data were chosen, how the input features were selected, and how the neural networks were trained. Finally the potential for the use of the new scheme within Nimrod is assessed.

3 Training and Validation Data

At the Met. Office, Meteosat imagery is processed through the Autosat system, which reprojects images into polar stereographic coordinates, and derives a range of products, including visible reflectivity and infrared brightness temperatures, from the instrument counts. Infrared grey-scales range from 4 to 251, and represent brightness temperatures of 198K to 308K respectively. The Nimrod nowcasting system ingests these two Meteosat products, corrects them for parallax and sun angle and maps them onto a 5km grid within the Nimrod domain (Golding, 1998).

The radar data come from a network of 15 radars sited around the British Isles, and each image undergoes processing including identification and removal of corrupt images; removal of spurious echoes (with the help of the Meteosat thresholding scheme); correction of echo intensity with range, bright band contamination and orographic enhancement below the radar beam; and rain gauge adjustment. The images are then composited into the Nimrod domain, at 5km resolution.

For the development of the PoP classifier, 48 sets of Meteosat and radar composite images were restored from archive, covering the period July 1995 to June 1997 (including visible reflectivity data when available). Two sets were restored for each month at random times, coinciding with half hourly Meteosat imagery (see table 2).

Timestamp	Visible	Non-missing	Raining	Timestamp	Visible	Non-missing	Raining
199507012230	N	36986	2137	199607010100	N	32073	3594
199507151230	Y	36137	1842	199607152030	N	32942	179
199508010830	Y	32207	1336	199608010600	N	33973	2974
199508152100	N	33124	61	199608150900	Y	32073	200
199509012300	N	35368	3587	199609010200	N	33791	1733
199509150430	N	33973	2797	199609151730	Y	33973	21
199510012100	N	30369	427	199610011630	Y	33973	4064
199510152330	N	33973	1788	199610152100	N	33973	4985
199511010930	Y	26402	630	199611011300	N	33973	3524
199511130600	N	33977	1653	199611150230	N	33973	1050
199512181230	Y	26550	1072	199612011730	N	33973	3349
199512251130	Y	34146	3041	199612150300	N	33973	1549
199601011930	N	36990	3589	199701011600	N	32073	2122
199601151130	Y	36956	1168	199701151500	N	37086	651
199602010500	N	33351	105	199702011300	Y	37086	2430
199602150930	Y	36986	1643	199702150030	N	37086	1723
199603012330	N	36986	225	199703011030	Y	33973	5981
199603150900	Y	35115	3160	199703152230	N	33973	5319
199604011230	Y	28469	1130	199704010630	N	37121	654
199604150130	N	33973	338	199704151630	Y	37086	213
199605032230	N	33973	610	199705010730	Y	35259	143
199605150500	N	33973	668	199705152030	N	33973	1834
199606022100	N	33973	515	199706010200	N	37086	154
199606151530	Y	33973	28	199706150100	N	33462	2422

Table 2: *Timestamps (YYYYMMDDhhmm) of data files restored from the Nimrod archive. Availability of complete Meteosat visible reflectivity data is shown, as well as numbers of non-missing pixels and pixels with a rain rate* $> \frac{1}{32} mmh^{-1}$ *in the radar composites.*

For each radar composite, two additional radar composites were chosen before and after the time of interest, to form a radar movieloop with 5 half-hourly frames. These movieloops were examined for anaprop, often seen best in this way as patches of shimmering radar echoes bearing little relation to real rain systems, and often related to orography. Areas suspected of being contaminated with anaprop were marked on a copy of each image, to avoid selection for the neural network training set.

A total of 3200 samples of 17x17 pixels (85x85km), were selected from the collocated infrared brightness temperature and radar rainrate fields, from uncontaminated regions of the images (see figure 1). Samples were labelled as either no rain ($\leq \frac{1}{32}$mmh^{-1}), or as rain ($> \frac{1}{32}$mmh^{-1}), depending on the central radar rainrate (centre labelling) or the average radar rainrate over the sample (average labelling).

Average labelling is less sensitive, but provides a better correlation between cloud brightness temperature statistics and rain or no rain. Any subsample could therefore be chosen, with its own centre or average label, if required. Of the 3200 samples, 2097 had no rain as the central label and 1103 had rain, giving a dry to wet ratio of 1:1.9. This compares to the dry to wet ratio used by Cheng *et al.* (1993) of 1:2.8.

From this set of 3200 samples, 1142 were selected for which uncorrupted visible reflectivity data were known to be available (see table 2). These data were restored from archive, and added to the training and validation sets. Of the 1142 samples, 749 had no rain as the central label and 393 had rain, giving the same dry to wet ratio as for the infrared only cases.

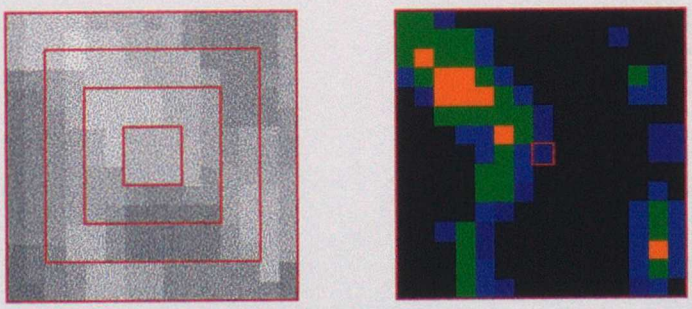

Figure 1. *An example of collocated Meteosat infrared brightness temperatures (white is colder), and radar rainrates (dark blue* $> \frac{1}{32}mmh^{-1}$*, light blue* $> \frac{1}{8}mmh^{-1}$*, green* $> \frac{1}{2}mmh^{-1}$ *and orange* $> 2mmh^{-1}$*; pixel size is 5km). Red boxes in the Meteosat sample represent the areas over which the candidate features were calculated, from 3x3 to 15x15 pixels.*

4 Feature Selection

Given samples of infrared brightness temperature or visible reflectivity pixels which we wish to correlate with a central or average rainrate label, it is possible to calculate numerous statistics or features over different sized regions (figure 1). The most obvious features are the infrared brightness temperature or visible reflectivity values at the central pixel; given this information, a neural network classifier should be able to classify as well as, or better than the threshold classifier discussed in Cheng *et al.* (1993), for example. However, it is possible to extract some simple local features, such as the mean, standard deviation, minimum, maximum, and the range and ratio of the maximum and minimum, providing a total of 7 features per spectral channel, including the central value.

It is also possible to extract textural features from regions of at least 3x3 pixels, such as grey-level difference vectors (Weszka *et al.*, 1976). A grey-level difference vector $\Delta g(\phi)$ is the absolute difference between two grey-levels with a fixed spatial relationship ϕ of angle, and distance (in pixels). In this work, four relationships were used: $\phi(0°, 1.00)$, $\phi(45°, 1.41)$, $\phi(90°, 1.00)$ and $\phi(135°, 1.41)$. Grey-level differences were calculated for every pair of pixels in the sample governed by these relationships, to produce a series of histograms $h_\phi(\Delta g)$. These histograms were then used to construct five statistics of interest: the mean, contrast, angular second moment, entropy and homogeneity (detailed in Pankiewicz, 1994). As an example, the homogeneity statistic H is given below:

$$\mathrm{H}(\phi) = \sum_{\Delta g=0}^{255} \frac{h_\phi(\Delta g)}{T} \cdot \frac{1}{1+|\Delta g|} \tag{1}$$

Here, T is the total number of grey-level differences in the sample. Mean and maximum values were calculated for the 4 spatial relationships described above, giving a total of 10 grey-level difference features per spectral channel.

To determine the ability of these 17 candidate features per spectral channel to discriminate rain from no rain, multidimensional Bhattacharyya distances were calculated (see Gu *et al.* (1991), for example), with mean vectors $\underline{\mu}$ and covariance matrices Σ obtained for both no rain (0) and rain (1) classes:

$$\begin{aligned} J_B &= \frac{1}{4} \cdot (\underline{\mu}_0 - \underline{\mu}_1)^T \{\Sigma_0 + \Sigma_1\}^{-1} (\underline{\mu}_0 - \underline{\mu}_1) \\ &\quad + \frac{1}{2} \cdot \ln \left\{ \frac{\left|\frac{1}{2}(\Sigma_0 + \Sigma_1)\right|}{\sqrt{|\Sigma_0||\Sigma_1|}} \right\} \end{aligned} \tag{2}$$

Bhattacharyya distances were calculated for sample sizes of 3x3, 7x7, 11x11 and 15x15 pixels, for both centre and average labels, for all possible combinations of feature vector for infrared features only (a total of ${}^{17}C_1 + {}^{17}C_2 + \ldots + {}^{17}C_{17} = 131071$, where ${}^nC_r$ is the number of combinations of r features from n). They were also calculated

for all possible combinations of feature vector up to 9 dimensions for infrared plus visible features (a total of ${}^{34}C_1 + {}^{34}C_2 + \ldots + {}^{34}C_9 = 77663191$ combinations). For each feature vector dimension, the 10 feature vectors with the largest Bhattacharyya distances were noted.

4.1 Infrared features

The results for infrared brightness temperature features alone are presented in figure 2, showing the Bhattacharyya distances for the most separable feature vectors. As the number of features used is increased, their ability to separate the classes increases, but at a decreasing rate. Also, there is more ability for textural features to discriminate average labelled samples. If the sample size increases, so does the ability to separate the classes, but again at a diminishing rate. In particular, centre labelled samples were unable to discriminate further for samples of 11x11 pixels or more. For an operational classifier, speed is important, and the smaller the sample size, and the fewer the textural features used, the better.

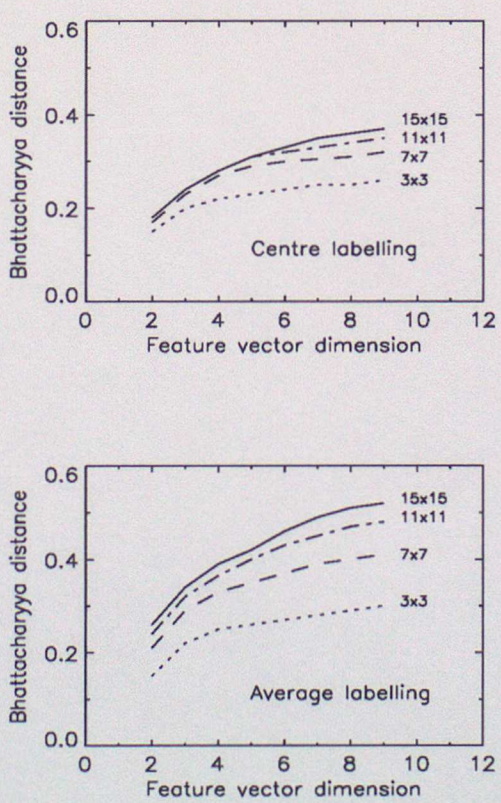

Figure 2. *The largest Bhattacharyya distances obtained for different sample sizes of infrared brightness temperature data only, for both centre and average labelled classes, as feature vector size increases from 2 to 9.*

This suggests that for centre labelled samples (more sensitive to the detection of anaprop), 7x7 pixels and 5 features are adequate (resulting in a Bhattacharyya distance of 0.29, compared to 0.37 for 15x15 pixel samples and 9 features). The corresponding feature vector consists of the minimum, maximum, range, ratio and central pixel value. Figure 3 shows these feature values (plus the most useful textural feature, the mean homogeneity), normalised by the means and standard deviations to provide values in the range 0 to 1, for the no rain and rain classes, for 1000 centre labelled samples per class.

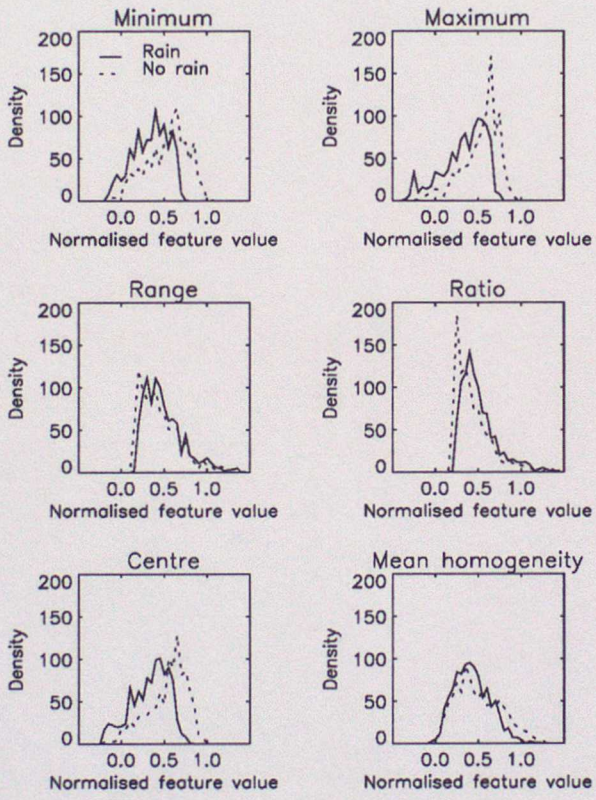

Figure 3. *The five feature distributions used to train the infrared only neural network, as well as mean homogeneity, normalised in the range 0 to 1, for 1000 samples per class, using centre labelling.*

4.2 Visible and infrared features

In order to speed up the analysis, Bhattacharyya distances were only calculated for sample sizes of 7x7 pixels, using centre labelling for cases where both visible reflectivity and infrared brightness temperature data were available. In this case, the

largest Bhattacharyya distances were found to continue increasing at a feature vector size of 9 dimensions, with textural features starting to be used at 9 dimensions.

The largest Bhattacharyya distance of a feature vector without textural features was found to consist of the minimum, maximum, range and ratio of the visible reflectivity, plus the minimum, maximum, range and ratio of the the infrared brightness temperature (8 dimensions). At 9 dimensions, the largest Bhattacharyya distance included the same features as at 8, together with the maximum visible reflectivity entropy measure of texture.

The best features selected to train the infrared, and the visible and infrared neural networks are listed in table 3.

INFRARED ONLY	VISIBLE AND INFRARED
Infrared minimum	Infrared minimum
Infrared maximum	Infrared maximum
Infrared range	Infrared range
Infrared ratio	Infrared ratio
Infrared centre	Visible minimum
	Visible maximum
	Visible range
	Visible ratio

Table 3: *Selected features used to train the infrared, and the visible and infrared neural networks.*

5 Neural Network Training

Feature selection showed that 5 or 8 features depending on availability of visible reflectivity data are most useful at discriminating 2 rain classes (no rain and rain with a threshold of $\frac{1}{32}\text{mmh}^{-1}$). Discrimination of a feature space into a number of classes is a typical statistical pattern recognition problem, which can be solved using a number of methods, including a multilayer perceptron neural network. MLP neural networks provide fast, robust and highly accurate pattern classifiers, and are being used more and more in the remote sensing community (Pankiewicz, 1995).

Conventional MLP networks consist of a number of input nodes (representing the features calculated in the image, denoted by x_i), one or two "hidden layers" of nodes (j) and a layer of output nodes (k), the latter representing the output classes of interest (two in this case). The nodes in each layer are connected to all nodes in the next layer, so that for a network with 6 input nodes, 8 hidden nodes and 2 output nodes, there are a total of 64 connection weights, each representing connection strengths between various nodes.

MLPs can be trained using the backpropagation learning algorithm (Rumelhart *et al.* 1986), where an error function is calculated at each output node, and is then backpropagated through the network. For a given training sample belonging to a specific class, the feature values x_i are calculated and are input to the network. Initially the connecting weight values (w_{ij} and w_{jk}) are set randomly, and a sigmoidal thresholding function is used as the feedforward transfer function. The activity values on nodes at the hidden layer j and then the output layer k are calculated:

$$a_j = \frac{1}{1 + e^{-\sigma \sum(w_{ij} x_i)}} \tag{3}$$

$$a_k = \frac{1}{1 + e^{-\sigma \sum(w_{jk} a_j)}} \tag{4}$$

where σ measures the spread of the thresholding function. The value of the error function at the output layer is then determined, and is related to the difference between the calculated output values a_k, and the true output values o_k. The simple difference $o_k - a_k$ is not used, because of the way in which the error is backpropagated through the network (see Bishop, 1995). The error functions used in this work (see Pankiewicz, 1994 for a derivation), are:

$$\delta_k = \sigma a_k (1 - a_k)(o_k - a_k) \tag{5}$$

$$\delta_j = \sigma a_j (1 - a_j) \sum_k (\delta_k w_{jk}) \tag{6}$$

Weights between layers k and j are then updated with a gradient descent algorithm, the technique being used to search for the minimum value of the error function:

$$w_{jk}(t+1) = w_{jk}(t) + \rho \delta_k a_j + \alpha [w_{jk}(t) - w_{jk}(t-1)] \tag{7}$$

Here, ρ is the learning rate, and α (not to be confused with the assigned PoP field values) is a momentum term used to carry forward previous weights changes. For weights between i and j, the same equation is used with j replaced by i, k replaced by j and a_j replaced by x_i. In this way, the error function decreases to a minimum after repeated presentation of training samples, each presentation being referred to as an epoch.

Numerous parameters affect the performance of such a network, including the number, type and normalisation of features calculated in the image samples, the number of hidden 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, the rate of adaption of weights ρ, and the momentum factor α.

5.1 Infrared features

The 3200 samples were split randomly into a training set and a validation set, containing two-thirds and one-third of the data respectively. An MLP network was

trained with the input features described above, as well as a bias node (a standard technique in which an extra input is included with x_0 set to a value of 1). Weight values were recorded every 5 epochs, and the validation set were used to obtain the average Probabilities of Detection and False Alarm Rates. The aim was to increase the PoD and decrease the FAR. Note however that to get the same ratio of wet to dry pixels, we require FAR = 1-PoD. If FAR > 1-PoD, then the scheme is overestimating the number of wet pixels. In tests, a 1:2.8 wet to dry distribution of samples was used in order to compare the results with the work of Cheng *et al.* (1993).

After a large number of experiments in which the number of hidden nodes was changed, as well as values of σ, ρ, and α, the best network was found to produce a PoD of 66±9%, a FAR of 57±5%, and an Equitable Threat Score of 17±6%. These are average scores in the sense that the validation set was constructed out of samples taken during various synoptic conditions. An idea of the variances was obtained by calculating PoD, FAR and ETS for subgroups of samples from the total validation set. A total of 8 hidden nodes provided this solution (see figure 4), and values of $\sigma = 1.0$, $\rho = 0.5$ and $\alpha = 0.05$ were used. The input features were normalized according to zero mean, 4σ variance, and all samples were 7x7 pixels with centre class labelling.

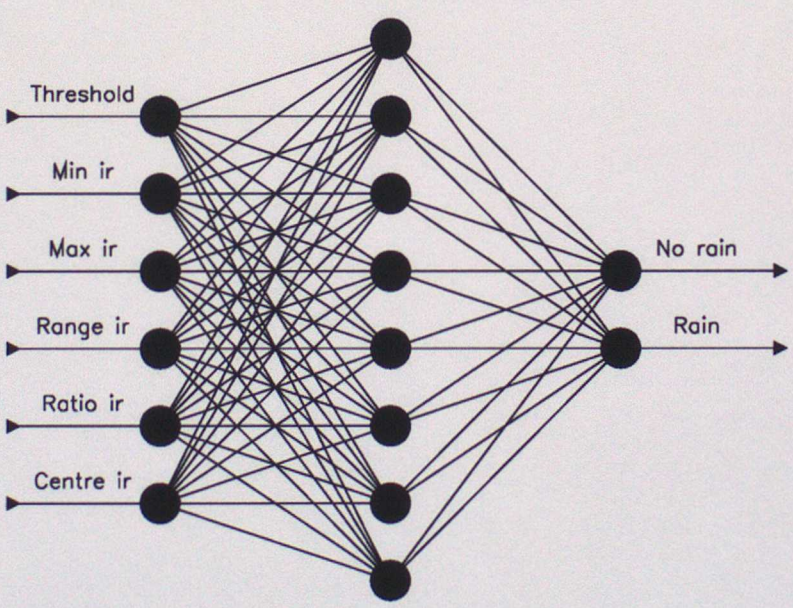

Figure 4. *The infrared neural network, showing the 5 input features (plus a standard bias input), 8 hidden nodes and 2 output classes.*

The results for the infrared only network can be compared with the best PoD of 61±24% and FAR of 47±12% for cold frontal cases obtained by Cheng *et al.* (1993), and their values of 28±21% and 74±11% respectively for cold-air convection, and 50±19% and 50±16% respectively for mesoscale convective systems.

An example of the resulting PoP where echoes were recorded in the Nimrod radar composite is shown in figure 5 for 20:30UT on 15th May 1997, together with the Meteosat infrared image and the radar composite. The radar composite was obtained after processing with the Meteosat threshold classifier, and a large number of spurious echoes still remain over Ireland, Strathclyde, Grampian region, the Midlands, the southeast and the Normandy and Brittany coast. The neural network PoP classifier has assigned PoP values of around 0.02-0.3 over Northern Ireland, Strathclyde and the Grampian region, 0.02-0.5 over the Midlands, 0.01-0.3 over the southeast and 0.02-0.3 over the Normandy and Brittany coast. However, PoP values of around 0.5 are found for the band of rain over the Firth of Forth, and 0.5-0.7 for the rain off the Lincolnshire coast.

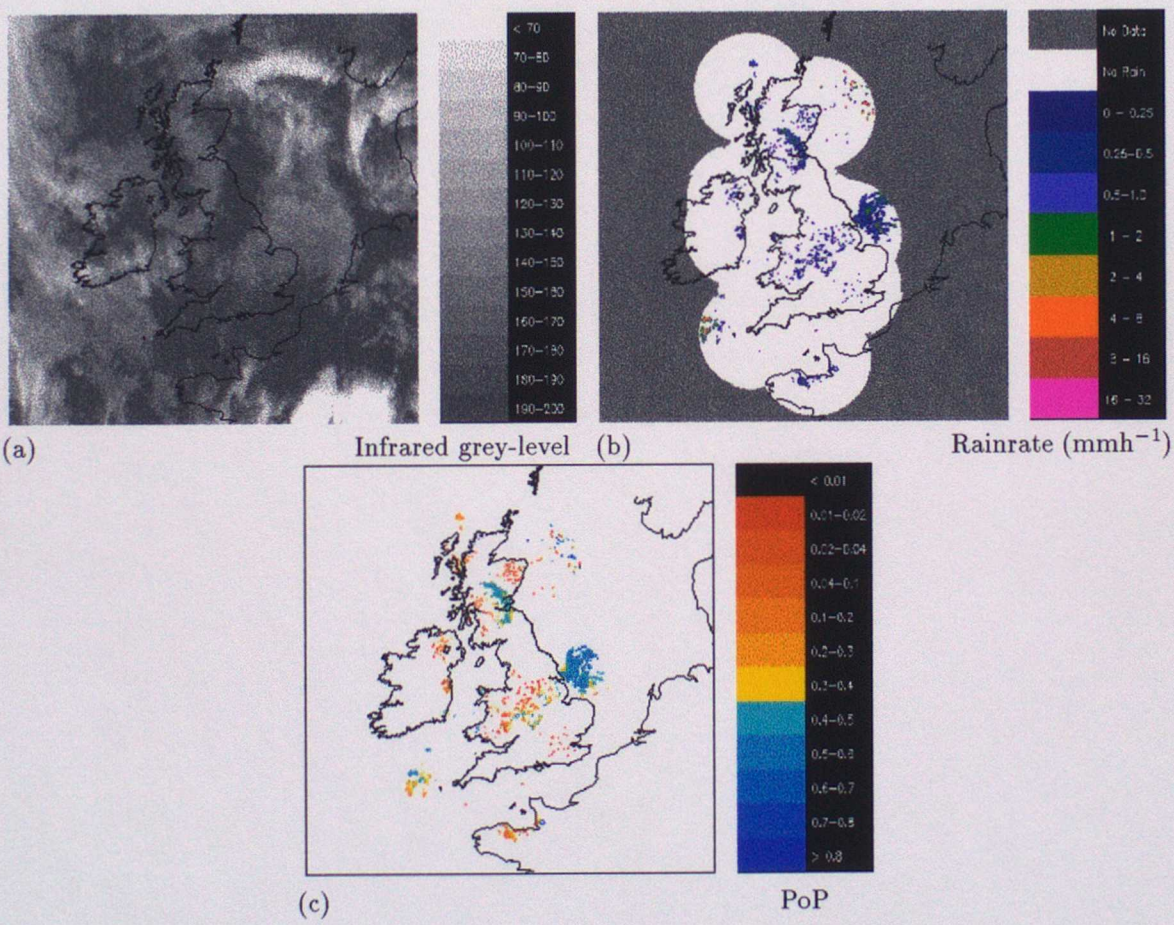

Figure 5. *An example of PoP values from the infrared feature neural network: (a) Meteosat infrared image for the Nimrod area at 5km resolution, at 20:30UTC on 15th May 1997, (b) the Nimrod radar composite, and (c) the neural network PoP field where radar echoes were recorded.*

5.2 Visible and infrared features

For the visible and infrared feature network, the 1142 available samples were again split into a training set from two-thirds of the data, and a validation set from the remaining third. The same procedure was adopted to search for the best classifier, and in this case, the best network was found to produce a PoD of 69±10%, a FAR of 41±3%, and an ETS of 33±7%, all superior scores to the infrared only PoP classifier. A total of 12 hidden nodes were required, together with training parameter values of $\sigma = 1.0$, $\rho = 0.2$ and $\alpha = 0.01$. All other neural network aspects were the same as for the network using infrared features alone.

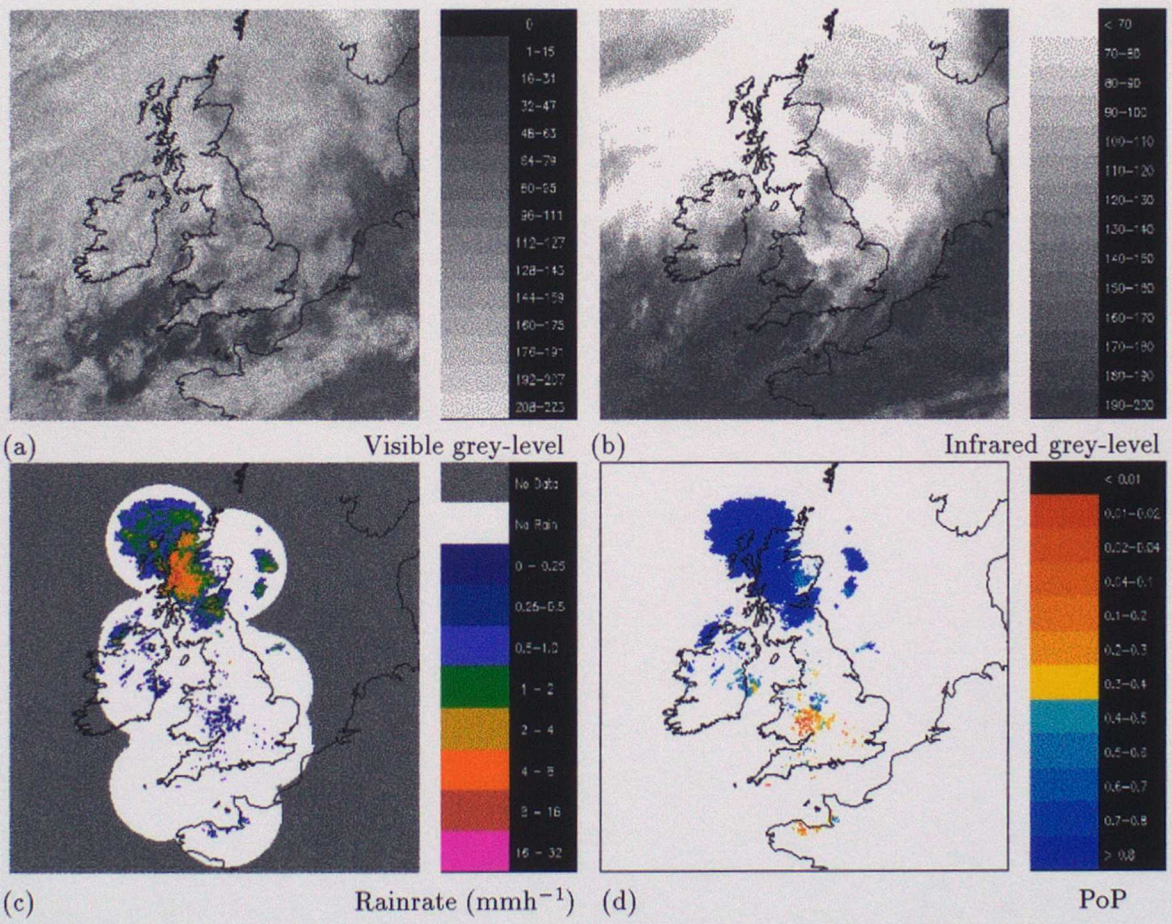

Figure 6. *An example of PoP values from the visible and infrared feature neural network: (a) Meteosat visible image for the Nimrod area at 5km resolution, at 10:30UTC on 1st March 1997, (b) Meteosat infrared, (c) the Nimrod radar composite, and (d) the neural network PoP field where radar echoes were recorded.*

Figure 6 shows an example of visible and infrared PoP values for 10:30UT on 1st March 1997. Again, the radar composite was obtained after processing with the Me-

teosat threshold classifier, and a large number of spurious echoes have been retained over the Welsh Borders and West Midlands, and the Normandy and Brittany coast. The neural network PoP classifier has assigned PoP values of around 0.02-0.4 over the Welsh Borders and 0.04-0.4 over the Normandy and Brittany coast, whilst retaining PoP values of 0.5-1.0 over the large areas of rain to the northwest of Ireland, to most of Scotland and out in the North Sea east of Aberdeen.

6 Nimrod Case Study Assessment

In the Nimrod anaprop removal scheme, actual PoP values are not used in the calculation of a final probability map. Alpha values are used instead, where in general, an alpha is defined as:

$$\alpha = \frac{P(W)}{P(D)} \tag{8}$$

where $P(W)$ is the probability of a pixel being wet and $P(D)$ is the probability of a pixel being dry. An alpha value can be said to be the "odds" that a pixel is actually wet and ranges from 0 (definitely no rain) through 1 (the climatological probability of precipitation) to large values up to 100 (definitely rain). The use of alpha values in the Nimrod anaprop removal scheme is described by Pamment and Conway (1998). The neural network PoP product needs to be converted into a field of alphas before it can be implemented into the current anaprop removal scheme.

The neural network PoP field must be shown to provide a better diagnosis of precipitating and non-precipitating cloud than the current field (described by Cheng *et al.*, 1993), if it is to be implemented into the Nimrod anaprop removal scheme. The current PoP field diagnoses shallow and small-scale convection poorly and this is particularly evident when only infrared data are available: the neural network classifier would be of particular value if it can show improvements in such conditions. The classifier must identify areas of non-precipitating cloud in order that spurious echoes in radar images can be effectively removed. However, an overriding requirement of the anaprop removal scheme is that precipitation echoes must not be removed, therefore, the classifier must also accurately identify areas of precipitation.

The performance of the new classifier was evaluated during two periods, between 28th May and 8th September 1998, and 28th October and 18th December 1998.

6.1 28th May to 8th September 1998

In this period, the assessment was designed to measure the performance of the neural network PoP classifier, without surface reports or short period forecasts. The PoP from the classifier was converted into a field of alphas (neural network alphas) which

could be directly compared to the Nimrod Meteosat field of alphas (Nimrod alphas) of the current anaprop removal scheme. In the current scheme, an alpha value of 1 is the threshold with values below 1 assumed to indicate dry conditions and any echoes falling in these regions being deleted. Alpha values of 1 or greater are assumed to indicate a climatological probability of rain or higher and any echoes falling in these regions are retained. A set of linear transforms were used to convert the neural network PoP into alpha values. These transforms were derived by examining a number of radar composites and finding a transformation which produced a balance of alpha values for wet and dry radar pixels in a similar way to the current Nimrod alphas field. The threshold at which echoes would be deleted or retained was derived by finding a value which produced a balance of anaprop removal and rain retention similar to the current Nimrod alphas. The same linear transforms were applied to the neural network PoP field, regardless of whether infrared only or visible and infrared data were used. In this period, the transforms used meant that, for the neural network PoP, the threshold probability below which echoes are removed was 0.143 (i.e. a neural network PoP of 0.143 is equal to an alpha of 1).

On an hourly basis, the Nimrod and neural network alphas were automatically compared to a Nimrod radar composite and a count made of the number of pixels which would have been deleted by each of the two fields of alphas. The assessment was carried out on a daily basis and cases where there were large differences between the number of pixels deleted by each method were examined. A subjective assessment was made as to the effectiveness of the removal of clutter and anaprop, and the retention of echoes associated with precipitation, using the criteria shown in table 4.

Over the summer assessment period (28th May - 8th September 1998), 141 cases were examined and scored, and the results are presented in tables 5 and 6. Note that the averages referred to in tables 5 and 6 are weighted by the scores in each case - the ideal values would be 1.00.

The neural network scheme performed better than the Nimrod scheme in terms of the accuracy of anaprop diagnosis. 97% of the cases examined showed the neural network scheme to have either diagnosed all cases of anaprop correctly, or retained only some light anaprop or clutter. This compares favourably with the same figure for the Nimrod scheme, which was 84%.

In terms of the accuracy of rain diagnosis, the current Nimrod scheme performed better than the neural network scheme. 77% of the cases examined showed the Nimrod scheme to have either diagnosed all cases of rain correctly or removed only an insignificant amount of light rain (e.g. from the edges of rain clouds). The neural network scheme diagnosed 60% of cases to this level of accuracy, with the remainder having rain deletion likely to give a misleading analysis and forecast, or deletion of significant amounts of rain.

Score	Description of anaprop assessment scores	Description of rain assessment scores
1	No misdiagnosis of anaprop	No misdiagnosis of rain
2	Some light anaprop retained	Some light rain removed - e.g. from edges of rain cloud (insignificant amount)
3	Anaprop retention likely to cause problems for TV applications, but does not significantly affect precipitation forecast	Deletion of rain likely to give misleading analysis and forecasts
4	Significant amount of anaprop retained	Significant amounts of rain deleted

Table 4: *Description of assessment scores.*

Scheme	1	2	3	4	*Average*
Nimrod alphas	50	68	20	3	*1.83*
Neural network alphas	87	50	4	0	*1.41*

Table 5: *Frequency of each score in the anaprop assessment.*

Scheme	1	2	3	4	*Average*
Nimrod alphas	84	25	21	11	*1.71*
Neural network alphas	53	32	28	28	*2.22*

Table 6: *Frequency of each score in the rain assessment.*

The assessment of the schemes over the summer highlighted a known problem of diagnosis in the current Nimrod alphas. When only infrared data are available, the Nimrod Meteosat alpha fields often contain large areas which have been set to the climatological PoP (an alpha of exactly 1). This means that other sources of data such as synoptic reports and forecast alphas could lower the combined alpha value below the threshold of 1. The current scheme is therefore rather too sensitive to information from other sources. It was noted during the assessment that the neural network scheme produced a more useful alphas field, giving a much higher PoP when precipitation was actually diagnosed.

Figure 7 illustrates a case where the current Nimrod alphas field shows little skill in diagnosing precipitation, only setting some of the precipitating cloud to an alpha of 100. The images (c) and (d) in figure 7 show only the alpha values of the Nimrod and neural network schemes where there are echoes in the radar composite. The neural network image shows high alpha values which correspond with areas of precipitation shown on the Nimrod composite image. The Nimrod alphas (in this case, derived from infrared data only) show alpha values much closer to the threshold, with precipitation echoes shown in a dark purple colour having an alpha value of 100 (the threshold) which will be retained, but given a lower threshold from other sources of evidence, could be deleted.

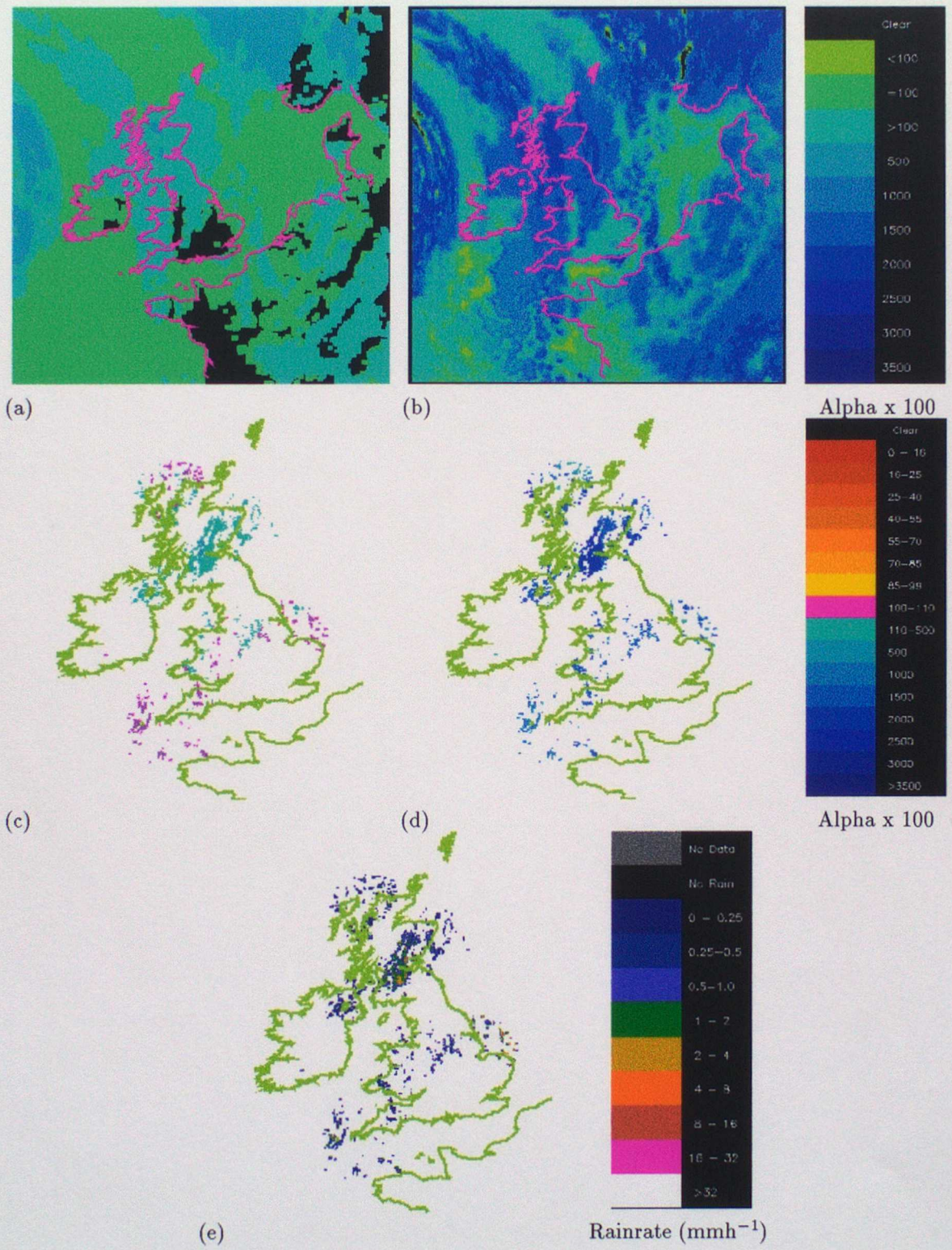

Figure 7. *A case showing the reduced skill of the Nimrod alphas when using infrared data only (00:00UT 11 February 1999): (a) Nimrod alphas, (b) neural network alphas, (c) Nimrod alphas where echoes are recorded in the Nimrod Radar Composite, (d) neural network alphas where echoes are recorded in the Nimrod Radar Composite, (e) Nimrod Radar Composite.*

To summarise, over the summer assessment period the neural network scheme was more useful at diagnosing areas of anaprop (i.e. identifying areas clear of rain) but there were a significant number of cases in which the rain diagnosis was worse than the Nimrod scheme. In practice, the neural network scheme may not delete many cases of rain because synop reports of rain will be used as an additional source of information, but it should perform to a similar standard of rain diagnosis as the current Nimrod scheme before implementation into the anaprop removal scheme.

6.2 28th October to 18th December 1998

The summer assessment established the potential of the new PoP classifier and some limitations but did not examine how effectively this information would combine with other sources of evidence (surface reports and short period forecasts). The aim of the winter assessment was to examine the performance of the new PoP classifier when implemented into the Nimrod anaprop removal scheme. This was carried out by running a full anaprop removal scheme with the new PoP classification implemented as an alphas field (i.e. replacing the current Nimrod alphas field), and combining it with the other sources of evidence. The number of pixels removed by each scheme (Nimrod and neural network) was recorded for each available radar site and cases of the largest pixel difference examined first. In this assessment, a score was also recorded for the amount of anaprop retained in the Nimrod "raw" radar images (before processing in the anaprop scheme).

Over the winter assessment period (28th October - 18th December 1998), 144 cases were examined, using the same scoring system as that used for the summer assessment. Results are presented in tables 7 and 8. Again, averages refer to averages weighted by the scores in each case.

Scheme	**1**	**2**	**3**	**4**	*Average*
Nimrod "raw"	10	19	37	78	*3.27*
Nimrod alphas	36	67	39	2	*2.05*
Neural network alphas	12	23	52	57	*3.07*

Table 7: *Frequency of each score in the anaprop assessment.*

Scheme	**1**	**2**	**3**	**4**	*Average*
Nimrod "raw"	144	0	0	0	*1.00*
Nimrod alphas	112	10	13	9	*1.44*
Neural network alphas	142	1	1	0	*1.02*

Table 8: *Frequency of each score in the rain assessment.*

The anaprop assessment scores show that the current Nimrod scheme removed more anaprop and clutter (i.e. the Nimrod scheme had more scores of 1 or 2) than the neural network scheme in 47% of cases examined. In 76% of cases examined, the neural network scheme retained a significant amount of anaprop or an amount likely to cause problems for TV applications. The neural network scheme performed better than the Nimrod scheme in terms of the amount of rain retained. In the rain assessment, the neural network scheme deleted significant amounts of rain in only 1% of cases, compared with 15% of cases for the Nimrod scheme.

6.3 Discussion of case study assessments

A direct comparison of the results obtained for 28th October - 18th December 1998 and those obtained for 28th May - 8th September 1998 cannot be made because the trials were comparing different products. However, some conclusions can be made about the differences between the performance of the schemes in the two assessments. In the summer period, both schemes performed well in terms of the amount of anaprop removed, with the new neural network scheme clearly performing better than the current Nimrod scheme. This is very different from the winter period, where the Nimrod and neural network schemes retained significant amounts of anaprop in a larger number of cases. The neural network scheme, however, performed significantly worse than the Nimrod scheme in the winter period. The rain assessments indicate that there has been an improvement in the amount of rain retained by both schemes, although this may be due to the use of synoptic reports.

Over the summer period, the neural network scheme performance was satisfactory in terms of anaprop removal, but would be unsuitable for implementation because of the amount of cases of significant rain deletion. It is because of these cases of rain deletion that a change was made on the 9th September 1998 to the way in which the neural network classifier PoP was converted into an alpha value. A new set of linear transforms was derived to be applied to the infrared only PoP field and a similar set of transforms maintained for when both the visible and infrared data were used to produce the neural network PoP. These changes shifted the threshold of rain deletion to be at a lower neural network PoP so that echoes would be retained at a lower PoP in the infrared only scheme. The new threshold meant that echoes would be retained if the infrared only PoP was greater than 0.02. This threshold was very low but it was found to produce suitable levels of anaprop and clutter deletion whilst retaining rain echoes, in a number of cases examined in the summer of 1998. The threshold was derived by finding a value which produced a balance of anaprop removal and rain retention similar to the current Nimrod alphas. This threshold was very sensitive and needed to be of an accuracy of two significant figures because changes of the order of 0.1 had a significant impact on the amount of pixels retained and removed.

Over the winter period, the results for the neural network scheme indicate that it would be unsuitable for implementation because of the poorer removal of anaprop

compared to the current scheme. One factor which is likely to have had a significant impact on the results over the winter period is the change in land and sea surface temperature. This might be causing a problem of contrast in the satellite imagery, with cold surface temperatures being misinterpreted as low cloud and hence resulting in relatively high estimates of PoP. Figure 8 illustrates the alpha fields for the Nimrod and neural network schemes and the radar composite and Neural Network PoP for a case on 21 January 1999.

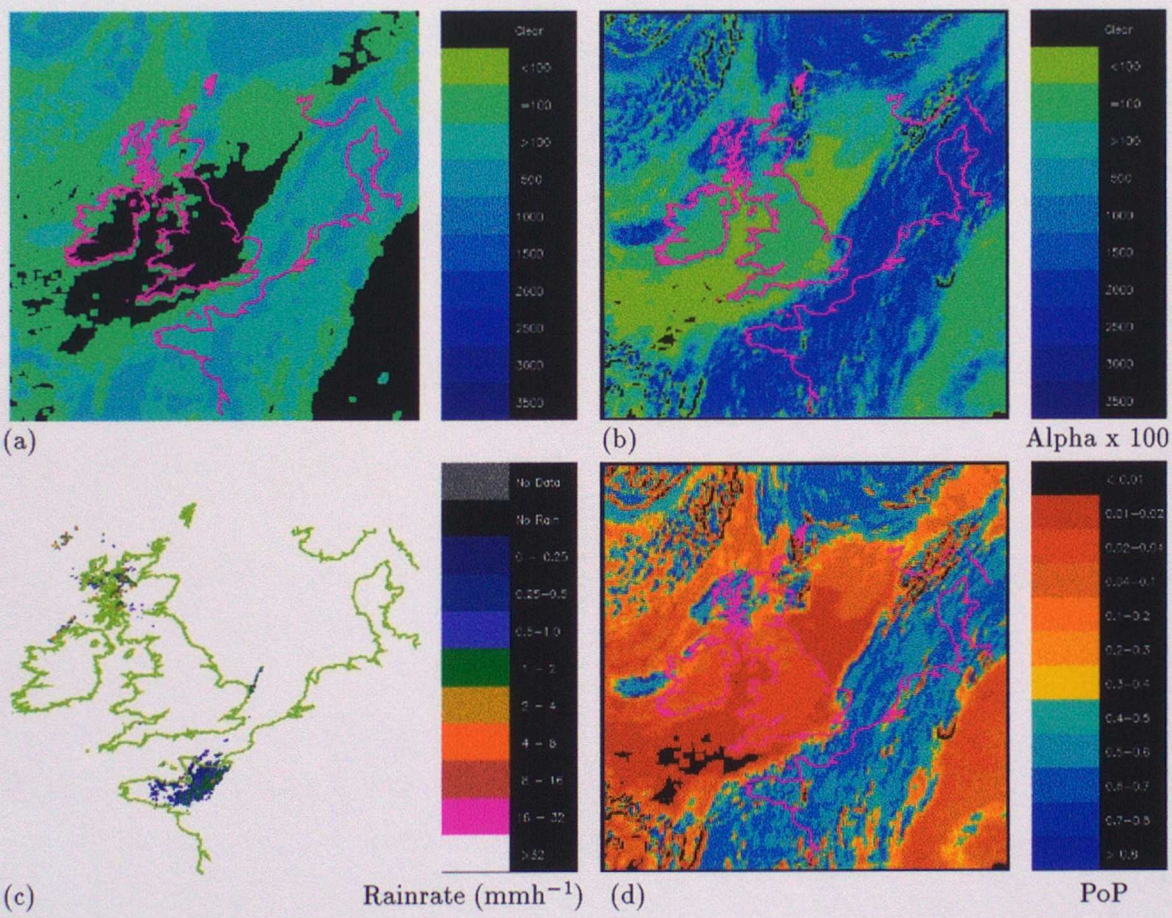

Figure 8. *A case study of 04:00UT 21 January 1999 showing higher PoP over the UK land mass in the neural network scheme: (a) Nimrod alphas, (b) neural network alphas, (c) Nimrod Rainrate Composite, (d) neural network PoP field.*

The neural network alphas show a higher PoP over the UK land mass compared with most of the surrounding sea areas. The Nimrod alphas are actually showing clear areas over the UK where the neural network alphas do not drop below an alpha of 1. The current Nimrod alphas scheme actually takes the surface temperature into account when calculating the PoP and it is likely that this must also be included in the new neural network PoP scheme for the classifier to perform effectively.

It is likely that the change to the way in which the PoP is converted into alphas has also had an impact on the amount of anaprop removal in the winter period. The deletion of too much rain over the summer period meant that a change was necessary to reduce the rain deletion, which may have resulted in a slight reduction in the amount of anaprop removed. It is very difficult to determine the exact impact that this change has had on the assessment results. The threshold at a PoP of 0.02 did produce suitable results in the summer cases examined but given the change in surface temperature in winter, this threshold change became significant because all PoP values were slightly raised. This meant that over the winter period, there were fewer cases when the neural network PoP actually dropped below 0.02.

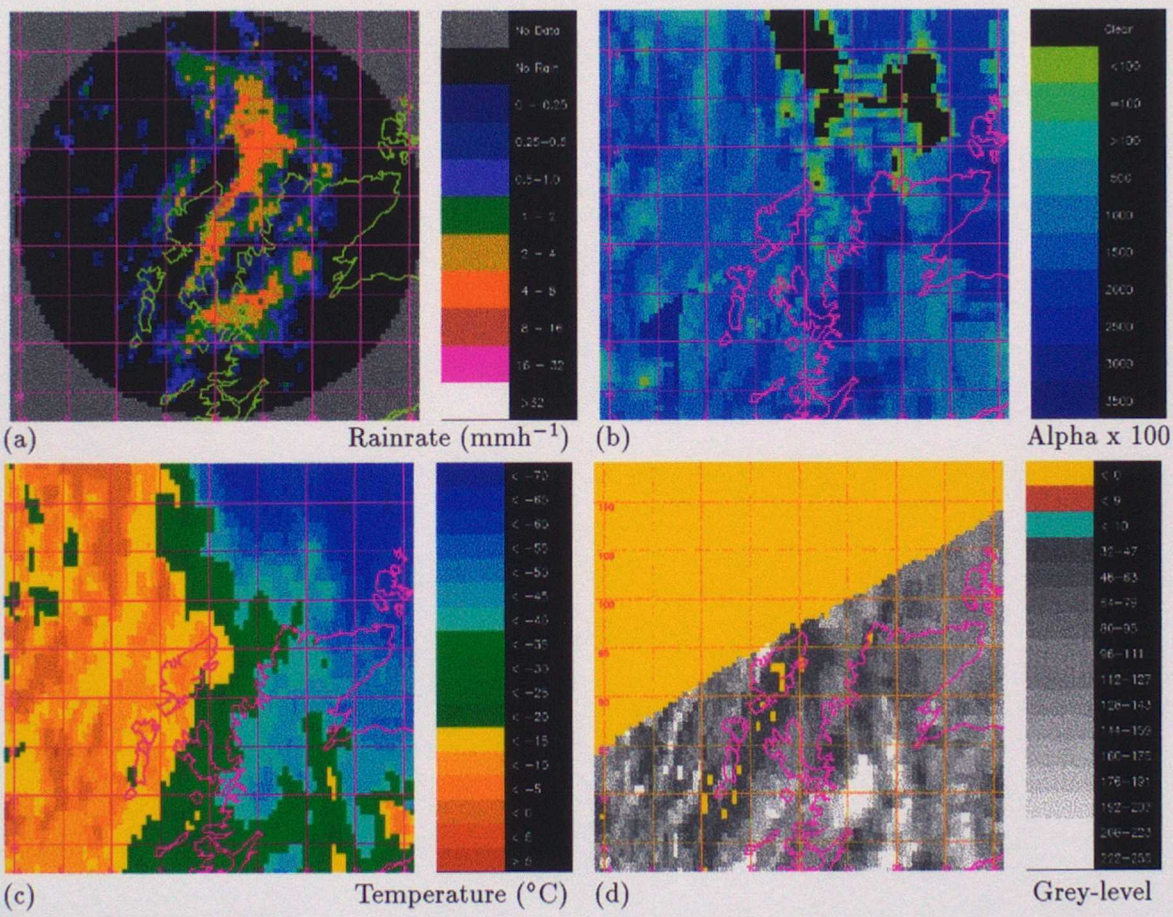

Figure 9. *Illustration of a case where significant amounts of rain were deleted at the edge of very cold cloud by the neural network scheme at 10:00UT 15th January 1999: (a) Nimrod Raw Radar image at Druima Starraig, (b) neural network alphas, (c) Meteosat infrared brightness temperature, (d) Meteosat visible reflectivity.*

It is also possible that the threshold of 0.02 set during the summer was based too strictly on the criteria that there should be no rain deletion. In practice this is very difficult to achieve and some cases of rain deletion should perhaps have been accepted

if it meant that more anaprop echoes could be removed. However, the success of the neural network scheme would then have been dependent on the value of other sources of data, i.e. if rain was not diagnosed by the neural network alphas, it would need to have been identified from another observational source.

During the assessment periods a further problem with the classifier was noted. Some cases were examined where the neural network PoP classifier had diagnosed regions of moderate to heavy rainfall as being of a very low PoP (i.e. diagnosed as clear in the neural network alphas). This problem was found to be infrequent but could result in serious rain deletion. Over the winter assessment period, this problem did not result in significant rain deletion because other sources of evidence gave high alpha values for the areas of precipitation. The cases of very low neural network PoP in areas of real precipitation appeared to occur when the infrared temperature of the cloud varied significantly over a short distance, for example, at the edge of extremely cold cloud. Figure 9 shows a case where significant amounts of rain were deleted at the edge of very cold cloud. This is unlikely to have arisen from timing errors (the Meteosat data are taken from the 10 minute B-format scans rather than the 25 minute full earth disk), and is more likely to be related to the way the network has learned to assign low PoP near cloud edges. In cases of extreme gradients, the network appears to assign very low PoP, which may not alway be true.

7 Conclusions and Recommendations

A Meteosat probability of precipitation classifier in the form of a neural network has been developed for use in the Nimrod anaprop removal scheme. The classifier can be run in one of two modes: when visible and infrared data are available or when infrared data alone are available. Trialling was performed using the classifier without other sources of evidence during summer 1998, and combined with other sources of evidence during winter 1998.

During the summer period, assessment of 141 cases indicated that the neural network classifier has the potential to provide an effective input into the Nimrod anaprop removal scheme, particularly in diagnosing shallow convection when only infrared imagery is available. Correct diagnosis of rain was not, however, as good as the current Nimrod scheme. As a result the method by which the PoP field was converted to an equivalent alpha value was amended.

Over the winter period, the neural network scheme showed a worse diagnosis of anaprop than the current Nimrod scheme. The main problem in diagnosing anaprop appeared to be related to surface temperature, where cold surface temperatures in cloud free conditions were assigned a relatively high PoP. This was likely to have been more apparent because of the reduced PoP threshold used in the winter period which became less suitable with the change in surface temperature. Another problem

was evident in both the summer and winter periods, related to a poor diagnosis of PoP where extreme gradients were observed in the satellite imagery. On the basis of the assessment results, the neural network PoP classifier is not currently suitable for operational implementation within Nimrod.

Perhaps the most effective method for removing spurious echoes would be to convert the British Isles radar network to Doppler systems, in which the velocities associated with a radar echo could be used to imply whether the echo is from a moving rain system, or spurious. However, there are no current plans to upgrade the radar network for at least 5 to 10 years.

The neural network scheme showed early promise, particularly for improving PoP diagnosis based on infrared data only. Although a number of problems emerged during the trial period, it became apparent that nearly all of these could be remedied. Given that few alternatives exist to improve the analysis of spurious echoes from independent data, and that the neural network technique can be refined with the use of surface temperature, snow cover and a surface climatological albedo field (all used in the current scheme), it is recommended that an improved neural network PoP classifier is developed, and trialled next winter. In addition, the current threshold used for infrared data only should be reviewed, and the effect of extremes of visible reflectivity and infrared brightness temperature (and their gradients) should be examined.

8 References

Bishop, C.M., 1995: Neural networks for pattern recognition. *Clarendon Press, Oxford*, 482pp.

Cheng, M., Brown, R. & Collier, C.G., 1993: Delineation of precipitation areas using Meteosat infrared and visible data in the region of the United Kingdom. *J. Appl. Meteor.*, **32**, 884-898.

Golding, B.W., 1998: Nimrod, a system for generating automated very short range forecasts. *Meteor. Appl.*, **5**, 1-16.

Gu, Z.Q, Duncan, C.N., Grant, P.M., Cowan, C.F.N., Renshaw, E. & Mugglestone, M.A., 1991: Textural and spectral features as an aid to cloud classification. *Int. J. Remote Sensing*, **12**, 953-968.

Johnson, C.J., 1998: A description of the methods used in the current Nimrod anaprop removal scheme. *Observation Based Products Technical Report no.10, UK Meteorological Office, Bracknell, UK,* 16pp.

Lovejoy, S & Austin, G.L., 1979: The delineation of rain areas from visible and IR satellite data for GATE and mid-latitudes. *Atmos. Ocean*, **20**, 77-92.

Pamment, J.A. and Conway, B.J., 1998: Objective identification of echoes due to anomalous propagation in weather radar data. *J. Atmos. and Ocean. Tech.*, **15**,

98-113.

Pankiewicz, G.S., 1994: The Application of Pattern Recognition Techniques to Cloud Detection. *Forecasting Research Division Technical Report no.129, UK Meteorological Office, Bracknell, UK,* 74pp.

Pankiewicz, G.S., 1995: Pattern recognition techniques for the identification of cloud and cloud systems. *Meteorol. Appl.*, **2**, 257-271.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. 1986: Learning representations by back-propagating errors. *Nature*, **323**, 533-536.

Weszka, J.S., Dyer, C.R. & Rosenfeld, A., 1976: A comparative study of texture measures for terrain classification. *IEEE Trans. System, Man and Cybernetics*, **6**, 269-285.