

Reproducing a Subjective Classification Scheme for Atmospheric Circulation Patterns over the United Kingdom using a Neural Network

G. C. Cawley¹ and S. R. Dorling²

¹ School of Information Systems,

² School of Environmental Sciences,
University of East Anglia, Norwich, U.K.

Abstract. Atmospheric circulation patterns are currently classified manually according to subjective schemes, such as the Lamb catalogue of circulation patterns centred on the United Kingdom. However, the sheer volume of data produced by General Circulation Models, used to investigate the effects of climatic change, makes this approach impractical for classifying predictions of the future climate. Furthermore, classification extending over long periods of time may require numerous authors, possibly introducing unwelcome discontinuities in the classification. This paper describes a neural classifier designed to reproduce the Lamb catalogue. Initial results indicate the neural classifier is able to out-perform the currently used rule-based system by a modest, but significant amount.

1 Introduction

Atmospheric circulation patterns can exhibit a great deal of day to day variability both in the lower layers of the troposphere and at so-called upper levels ($\approx 5\text{km}$ above ground level). Analysis shows however, that there are a finite number of such patterns which neatly summarise the range of this variability; each circulation pattern for a given day can be allocated to one such class. Subjective schemes have been developed for the United Kingdom and Europe, extending back to the nineteenth century on a daily basis [7, 2]. These schemes are known as the Lamb catalogue and the Grosswetterlagen respectively. While criticism can be made of any subjective index, Principle Component Analysis [6] confirms that, for example, the Lamb catalogue does provide a very useful indicator of changes in circulation both on the day to day scale and on the climatic time-scale. Furthermore, the Lamb index has been utilised in numerous studies relating atmospheric circulation to a suite of environmental problems (e.g. [1]).

Atmospheric scientists are increasingly interested in predictions of atmospheric circulation made by General Circulation Models (GCMs). To subjectively classify the profusion of such output now available would be impossible; an automated approach is required. GCMs are observed to reproduce today's climate more satisfactorily in terms of circulation patterns than temperature and precipitation patterns. Since a strong relationship exists between circulation on the

regional scale and local temperature and precipitation conditions, some studies are using predicted changes in circulation patterns to *down-scale* what those circulation changes mean for local temperature and rainfall changes [3]. Subjective classifications may suffer from discontinuities in the record. Such discontinuities may have a number of sources, such as a change in the person carrying out the classification or a change in the raw data used for the purposes of the classification. The availability of upper air data, starting around the time of the Second World War, may represent such a discontinuity in both the Lamb Catalogue and Grosswetterlagen.

In this first study, we aim to implement an objective neural classifier designed to replicate, as closely as possible, the subjective Lamb catalogue. Being an automatic method, it is well suited to classifying the output of general circulation models in the study of climate change. The added benefit of a system designed to closely reproduce a subjective classification scheme, in this case the Lamb catalogue, would be to allow the existing catalogue to be continued seamlessly for an indefinite period. Here we report on the development of the neural classifier and compare the ability of the net to reproduce the Lamb classification on a day to day basis against an objective algorithm known as the Jenkinson scheme [4] as utilised by Jones *et. al.* [5].

2 The Lamb Catalogue

The Lamb catalogue describes twenty-seven distinct atmospheric circulation patterns centred on the United Kingdom, which can be:

- Purely directional from a given point of the compass (N, NE, E, SE, S, SW, W, NW)
- Non-directional but described by an Anticyclonic or Cyclonic circulation pattern (A, C)
- A hybrid of the above two types (AN, ANE, AE, ASE, AS, ASW, AW, ANW, CN, CNE, CE, CSE, CS, CSW, CW, CNW)
- Unclassified (U)

The classification is performed on the basis of sea-level pressure and 500mBar height charts for Northern Europe, taken at midnight and midday, with reference also to the charts for the preceding days. As a result, a given day might be left as unclassified, either because the circulation pattern for that day was weak or poorly defined, or because a transition between two distinct circulation patterns occurred during the day. The record shows that some types are considerably more common than others and that all exhibit inter-annual variations in their frequency.

3 Methodology

The neural weather pattern classifier consists of a single multi-layer perceptron network. Each input neuron monitors the sea-level atmospheric pressure measured twice daily at sites arranged over a grid with a resolution of 5 degrees

latitude and 10 degrees longitude, as shown in figure 1. Before training, the input data is linearly scaled so that it has a zero mean and unit variance to simplify weight initialisation. The hidden and output units incorporate the standard symmetric logistic activation function. Each output unit corresponds to one of the seven major Lamb types: anticyclonic (A), cyclonic (C), northerly (N), easterly (E), southerly (S), westerly (W) and north-westerly (NW). Hybrid types, such as anticyclonic northerly, are represented by a combination of output units, in this case A and N. The north-westerly patterns though considered sufficiently distinct to merit a major type, are represented by the activation of three directional units N, W and NW as it was felt that this may simplify the learning task of the N and W output units. The target values for each output unit takes on values of -0.45 and 0.45 . This prevents the pointless growth of output layer weights in a futile attempt to generate an asymptotically high output value.

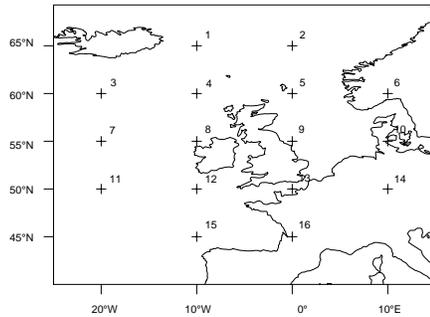


Fig. 1. Location of grid points over the British Isles. From [5]

The Rprop training algorithm used in this work, has been found to be both efficient and reliable for this application, and is also very tolerant of the adaptive learning parameters. For a complete description of the Rprop algorithm, see Riedmiller and Braun [8]. Fifteen years-worth of manually classified pressure data was divided randomly into three sections: training data (60%), validation data (20%) and test data (20%). To prevent over-fitting of the training data, the network configuration is saved when optimum performance on the validation data is reached. The statistically pure test data, that has not influenced the training procedure in any way, is then used to evaluate the network's performance.

4 Results

Figure 2 shows a graph of RMS error and the proportion of correctly classified patterns against cycles trained for a network with sixty-four hidden layer units. The network reaches a minimum RMS error on the validation data after 1000 epochs, at which time the network classifies 49.57% of the test patterns

correctly. This compares favourably with the 43.6% scored by the rule based system currently used [5].

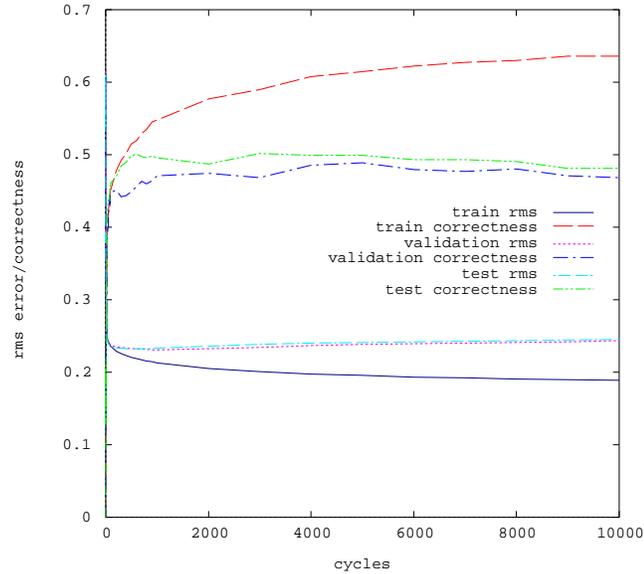


Fig. 2. Graph of RMS error against cycles trained for a multi-layer perceptron network with 64 hidden layer neurons.

Networks with 8, 16, 24, 32, 40, 48, 56 and 64 hidden units arranged in a single layer, were trained to determine the effect of network size on performance, measured by RMS error and the proportion of correctly classified patterns. In addition, a single layer perceptron network was trained to determine the extent to which the learning task is linearly separable. The results are depicted in figure 3. It is clear that the provision of hidden layer neurons immediately results in a substantial improvement in performance, but the improvement increases only slowly with increasing hidden layer size.

A confusion matrix for a neural Lamb classifier, for all available data from January 1973 to December 1988, is shown in figure 4. This provides a breakdown of the correctly classified patterns and false positive and false negative classifications for each of the 27 Lamb sub-types.

As noted in section 2, a given day may be left unclassified if the circulation pattern changed significantly during the day, as the neural classification is based on a single sea-level pressure chart, the classifier is unable to detect this. This accounts for a substantial portion of the false negative “unclassified” decisions. It can also be seen that the classifier is able to identify pure cyclonic and anti-cyclonic and pure directional types with a reasonable degree of accuracy, but performs less well on hybrid types. The wide dynamic range of the prior proba-

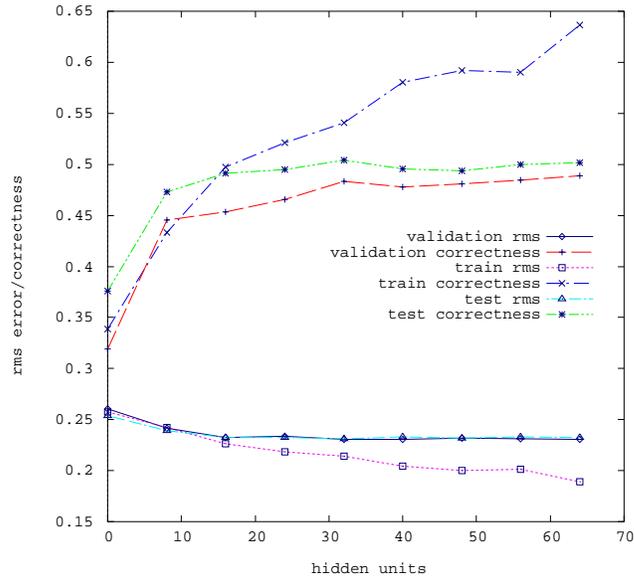


Fig. 3. Graph of RMS error and proportion of correctly classified patterns against hidden layer size, for training, validation and test data sets.

bilities for each class, is clearly a potential source of difficulty for any statistical classifier, including a neural network; for instance, westerly circulation patterns are particularly common due to the predisposition towards high pressure in the south and low pressure in the north, because of the temperature differential, whereas north-easterly patterns occur far less frequently. As a result, the overall error is minimised more efficiently by concentrating on the classes with the highest prior probability. As a result, the network correctly classifies 69.7% of the 726 westerly circulation patterns, but only 26% of the 61 north-easterly patterns.

5 Conclusions

In this initial study, a neural classifier has been constructed that is able to reproduce the Lamb catalogue more closely than the current rule based system by a modest, but significant amount. Both systems classify the prevailing circulation pattern according to the sea-level pressure at the same sixteen grid points centred on the U.K., since the subjective classification is also based on upper air patterns for each day, as well as the evolution of the circulation over the preceding days, this is clearly a simplified approach. We plan to take account of these other factors in future refinements of this work.

