Paper curl prediction and control using neural networks

PETER J. EDWARDS*, ALAN F. MURRAY, GEORGIOS PAPADOPOULOS, A. ROBIN WALLACE, JOHN BARNARD, AND GORDON SMITH

ABSTRACT

This paper describes the application of artificial neural network techniques to the papermaking industry, particularly for the prediction of paper curl. Paper curl is a common problem and can only be measured reliably off-line, after the paper is manufactured. By providing the machine operator with a tool that can predict curl, adjustments can be made to the machine and paper characteristics to maximize yield and reduce waste. Therefore, from characteristics of the current reel, we predict (a) whether the paper curl will be acceptable and (b) the level of curl. For both cases, we include confidence intervals. The techniques described in this paper are widely applicable to an industry where direct prediction of a quality measure and its acceptability are desirable, with a clear indication of prediction confidence.

Application:

This paper provides a methodology for applying neural network techniques to the task of predicting quality measures. Such an approach is appropriate when exact deterministic physical models are not available, while data examples of plant settings and the resultant quality are.

IN THIS PAPER, WE PRESENT AN approach to quality prediction in the context of the papermaking process, where deterministic models are unlikely to be available. Clearly, if a deterministic model exists that predicts a particular quality metric as a function of plant settings, that model should be used. Where rules are known that allow plant settings to be optimized, these should be used for decision-making or to provide automated plant control. In most examples of paper quality measures, rules and deterministic models are either nonexistent or incomplete. Artificial neural networks and multivariate statistical analysis provide means whereby data from previous plant operation may be used to build a "model," based not on detailed knowledge but on past experience. As this paper will show, any expert knowledge that exists should be included in the model-building process, but the model is fundamentally data-derived and is not the product of expert knowledge.

We describe the application of artificial neural network (ANN) techniques to the papermaking industry, particularly for the prediction of paper curl. Put simply, ANNs are nonlinear modeling tools that can be applied to tasks involving nonlinear data. An example of such a task is function modeling, whether that is a discriminant function in a classification task or a regression function in a process modeling task *(1)*. Such function modelers can be used for prediction problems. In practice, plant operators generally want three separate predictions:

- Is this roll of paper likely to be out of specification?
- If so, how badly is it out of specification?
- If I change some variable setting, will this out-of-spec. condition get better or worse?

We present parameters that characterize the current paper reel as inputs to a neural network and build a model so that a prediction can be made regarding whether the resulting level of curl will be within a required specification (i.e., in specification) or not—a classification task. In parallel, we present these same data to another network and form a model to predict the absolute level of curl, i.e., a regression task. Perhaps most importantly, we also put these two predictions in context by including confidence measures at every stage, thus providing the machine operator with a powerful and insightful tool. The machine operator is then presented with a red-light/green-light indication of paper acceptability, a neural regression model in which the parameters can be altered to reduce curl if necessary and a clear indicator of the reliability of both diagnostics.

Paper curl is a long-standing problem that has received much interest in papermaking research, where the causes of paper curl have been studied *(2)* and where attempts have been made to control curl using heuristic techniques *(3, 4)*. Paper curl reduces the quality of paper and is detrimental to the papermaking process at every stage of manufacture. In addition, it affects the customer, being the most common cause of sheet-feeding problems in nonimpact printing processes such as photocopying and laser printing *(5)*.

Curl is simply the tendency of paper to depart from a flat form and is affected by a number of complex, interrelated factors, including differing drying rates on the two sides of the paper during manufacture, relative variation in humidity or

*Correspondence may be sent via e-mail to pje@ee.ed.ac.uk.

moisture within the sheet, and mechanical stresses within the fibers. Additionally, a number of different modes of curl are recognized, e.g., inherent curl and curl occurring as a result of changes in the surrounding environment.

Traditionally, paper curl has been controlled with limited success using variation in the dryer temperatures or humidity levels. However, because it may only be measured offline after an entire roll has been produced, control of curl is difficult and costly. Curl is therefore a significant problem to the papermaking industry. Although out-of-specification paper may be repulped and recycled in limited quantities, significant curl wastes plant time, engineering time, and energy. Our aim is to develop a reliable model that can be used in decision support to render such waste "virtual," producing out-of-specification paper only in simulation. This, of course, is an ambitious goal.

As previously mentioned, ANNs are nonlinear modeling techniques that can be applied to the problem of predicting paper curl. These techniques are essentially a means of mapping a set of input—output vectors that have some nonlinear relationship. This is carried out using a set of internal parameters, or weights, that are adjusted, during "training," to encapsulate the relationship underlying the data vectors. In the case of paper curl, if there is a deterministic process causing the paper to curl and if we can select a number of parameters describing a reel of paper and its production that are related to this process, then by presenting examples of these parameters and the resulting curl to the network and adapting the internal parameters, we can develop a model. A more in-depth and rigorous discussion is provided in the literature *(1).* Importantly for this work, in a previous study *(6),* Myles showed that there is a nonlinear relationship within the paper curl process and that more conventional linear modeling techniques provide only limited solutions. Therefore, in our research, we use a multilayer perceptron (MLP) neural network for the modeling task.

Neural networks, therefore, may provide a solution to the difficult problem of predicting paper curl. The advantage of the neural network approach is that the user is not required to know the relationship underlying the data to solve the problem. The neural network can effectively encapsulate that relationship during the training process in its internal parameters. The disadvantage of the approach is related, in that there is the perception that at the end of training, the user still does not know what the relationship is. That is, the neural network solution is perceived as an impenetrable "black box."

In this paper, we seek to address this issue by maximizing the information available to the operator. It is already clear that there is no simple solution to the problem of paper curl. While the neural network may provide a solution, it will never be a simple one and it may only be a partial one. By providing the machine operator with (a) predictions of whether curl will be in specification, (b) an indication of the level of that curl, and most importantly, (c) a measure of how confident the network is about the predictions, the operator can judge when to make adjustments to the process and can know when the neural network cannot provide an accurate prediction.

In this paper, we describe the development of neural network models to encapsulate the nonlinear processes underlying paper curl. The results show that such models are able to reliably predict whether a roll of paper is likely to exhibit curl at a level less than or greater than some specified value, from parameters characterizing a reel of paper. In addition, models can be developed that are able to predict the absolute level of curl, with a sufficient level of accuracy to be useful in practice. These models can be used as powerful tools for reducing paper curl, thus enhancing quality and reducing waste.

In the following sections, we describe the database as well as techniques that we have used in preprocessing and training the neural network models. Results are then presented and discussed. In all of the work described, in terms of processing and modeling, the approach that we take can be applied to any task involving neural networks. Furthermore, these networks have been developed in the context of an imperfect data-collection process, which is typical of that found in many manufacturing operations; the combination of techniques used has particular relevance to such an environment.

THE DATABASE

The database provided by Tullis Russell for the purpose of the work described here has a number of limitations, including missing records and measurement errors. Not least in this respect is the measurement of curl itself. Although curl is a simple quality measure, measuring curl is far from trivial. While it would seem naturally advantageous to measure curl continuously, to date this has proven to be impossible, as standard techniques require the paper to be dried under controlled conditions before measurement *(4).* In addition, removing paper during manufacture is difficult due to the continuous nature of the task (e.g., dryer timing is related to machine speed) and risky due to the danger of a paper break during the process.

Therefore, at Tullis Russell, curl is measured after individual reels have been manufactured, leading to out-of-specification reels being scrapped (eventually to be recycled as broke), and a retrospective adjustment is made to machine settings according to unwritten heuristic rules developed by the skilled operators. The curl measurement is made using a sample of paper taken from the end of a reel and by cutting a cross shape in this sample using a template. A glancing-angle light source is then used to cast a shadow on the curling paper at the center of the cross. After a few minutes have lapsed to allow the paper to relax, the shadow is measured by hand and quantified to 5-mm intervals.

Unfortunately, there may be errors in the measurement due to quantification, operator error, paper misplacement, or failure to allow for sufficient relaxation time. Variability in the accuracy of curl measurement could lead to significant model error *(7),* and while we are developing an improved curl-measurement system, for the study reported in this paper, the limitations of the database are taken as an additional constraint on the modeling process.

Various parameters are measured prior to and during the manufacture of a reel of paper. A previous study *(6)* considered which parameters could be used to estimate curl. In our study, the choice was made pragmatically, considering data availability. The effectiveness of parameters in curl prediction was taken as a secondary issue. The final selection of parameters was as follows:

1. Softwood, %
2. Broke, %
3. Power used to process hardwood—refiner 1, kW
4. Power used to process hardwood—refiner 2, kW
5. Caliper (paper thickness) at the paper machine, µm
6. Ash content, %
7. Moisture content at the paper machine, %
8. Porosity measurement, mL/min
9. Difference between surface moisture levels after top-coat application, %
10. Paper grade (one of three).

These parameters were used to (a) classify whether the current process settings and paper specification would lead to curl that was within a required specification and (b) indicate the level of curl that would result. The preprocessing of these data is described as follows.

PREPROCESSING AND TRAINING

The task of training a neural network is not as straightforward as simply presenting the raw data to the network and obtaining a solution. In a companion paper *(8),* we describe in detail our approach to preprocessing the Tullis Russell raw data and training the neural network models. Here, we merely summarize that work. First, there is a need to combine the symbolic (e.g., paper-grade) and real data elements in a way that avoids putting any undue weighting on an individual parameter. Second, nonlinear optimization schemes, used for the purpose of training the neural network, perform dramatically better if colinearity in the data used for training is first minimized. A common technique for doing this is to extract the principal components of the data and to use these components to train the neural network models *(1).* The purpose of preprocessing the data is therefore to encapsulate all the information in the raw data set and to transform the information into a form that will maximize the potential of the neural network models.

Our approach to training the network uses state-of-the-art theory, where the networks, data, and predictions are all incorporated into a probabilistic framework. As such, network predictions are part of a probability distribution, where the prediction is the mean and there is an associated variance *(9).* This approach is important as the predictions, whether they be of actual curl levels or a probability of curl being out of specification, can be qualified with a confidence measure related to the variance of the distribution.

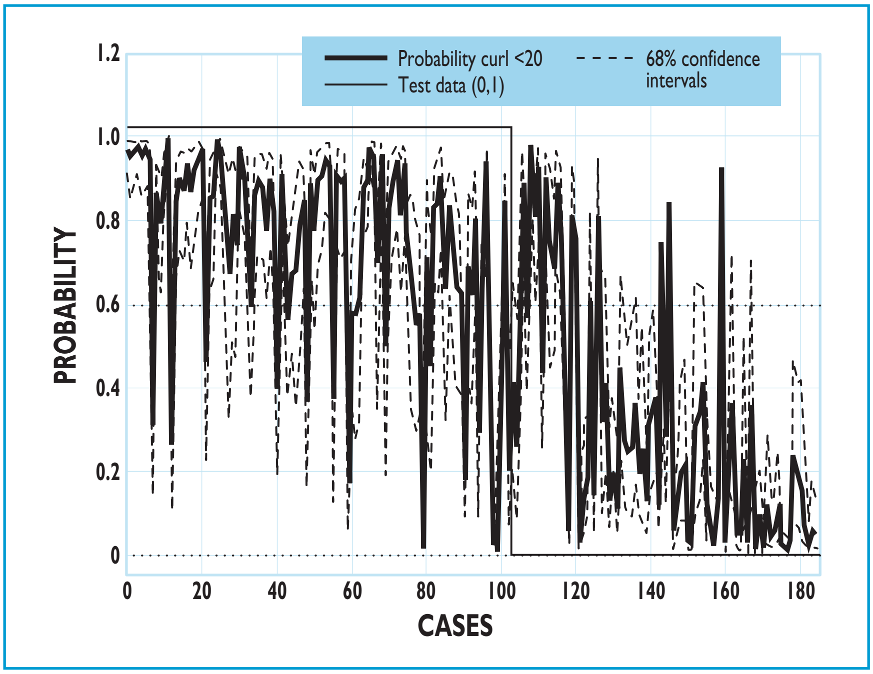
We will later return to the issue of confidence measures, noting that, as previously mentioned, one of the criticisms leveled at ANNs is that they can be seen as black boxes; having a measure of prediction reliability can reduce this perception. In addition, rather than using a single model for the classification and one for the regression task, here we use a committee of networks for each and combine the outputs to improve performance *(1).* In terms of preprocessing and training, therefore, a number of steps are carried out to achieve the modeling task. In the following section, results of experiments carried out using the trained networks are presented.

RESULTS

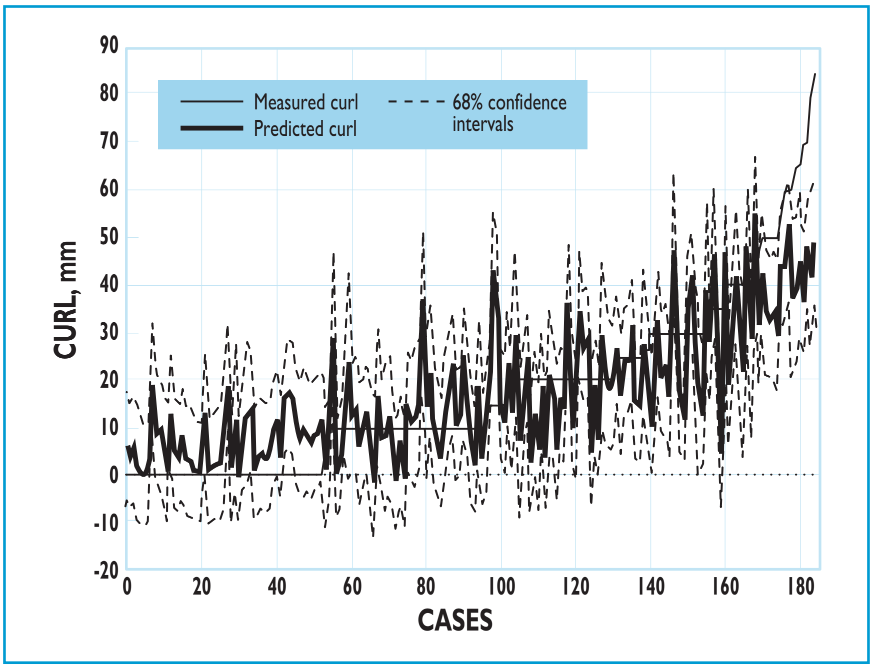
This section describes the results obtained from the neural network models that were trained to classify the current state of the manufacturing process as leading to out-of-specification paper in terms of curl and to predict the absolute level of that curl. All the results shown are test results, i.e., the performance of the models on previously unseen data.

Out-of-specification classification

In this experiment, 40 networks were trained to classify the current paper characteristics as leading to in-specification or out-of-specification paper. For the purpose of this experiment, a level of curl less than 20 was used as the limit of acceptance. This level was chosen as typical. Different grades of paper have different



1. Variation in the predicted probability of the curl being less than 20 for all the cases in the test data set. The test data are sorted for increasing levels of curl.



2. Variation in the absolute value of curl as predicted by the committee of networks and additionally as measured at Tullis Russell and contained in the test data set. The test data are sorted for increasing levels of curl.

acceptability levels. The results of these experiments are shown in **Fig. 1,** which depicts a classification error rate of 18.92%. The test cases are sorted into increasing levels of curl, and Fig. 1 shows that the majority of the errors (where the classification boundary is set at a probability of 0.5) occur at the center of the graph, i.e., for cases where the measured curl is 20 or near 20. The graph also shows 68% confidence intervals. Clearly the networks have been able to classify the cases to a usable degree of accuracy.

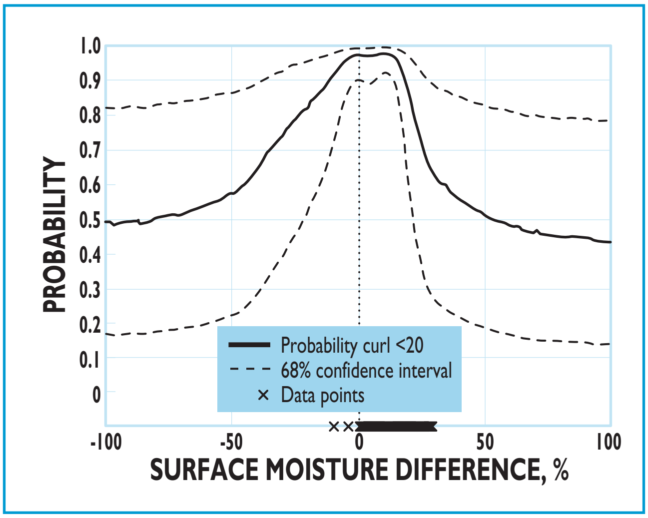
Curl prediction

In the previous section, the classification results were presented. **Figure 2** shows the results for 40 networks that were trained to predict the absolute level of curl and were tested on the same data set as that shown in Fig. 1. Clearly the model has encapsulated the trend underlying these measured data, although with some imprecision. The prediction of extremely high levels of measured curl is poor. This is perhaps due to this part of the model being underrepresented in the database (note that the data are densest for low curl) or the fact that the process is different for levels of curl greater than 40. In practice, however, some accuracy in the critical 10 < curl < 30 region is most important, and the predictor achieves adequate accuracy in that critical regime.

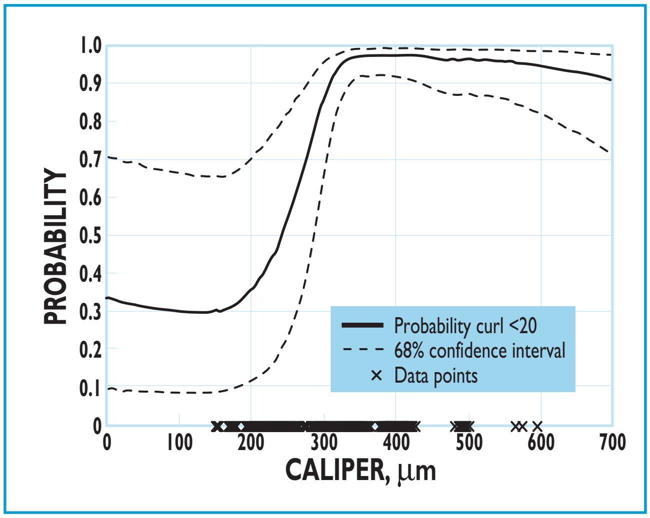
The prediction of curl is therefore possible to limited but usable accuracy using the database provided and a neural network model. In addition to the curl prediction, Fig. 2 also shows 68% confidence intervals for those predictions. These intervals are discussed in the following section.

CONFIDENCE MEASURES

In the previous section, results were presented showing that paper curl can be modeled using neural network techniques to a useful accuracy. In the experiments, test data were used to assess the models. While the data were previously unseen by the model during training, they were however taken from the same



3. Variation in the predicted probability of curl being less than 20 as the surface moisture variable changes. Also shown are the data points in that dimension and 68% confidence intervals.



4. Variation in the predicted probability of curl being less than 20 as the caliper variable changes. Also shown are the data points in that dimension and 68% confidence intervals.

database as the training data. In a real-world situation, data presented to the model by the operator will represent a further, and perhaps more difficult, test, where it is impossible to have a quantitative measure of prediction accuracy without making the paper and measuring curl in the usual way. Therefore, at this stage, we require a measure of how confident we are in any prediction to prevent the operator from having to use the predictions blindly as if from a black box.

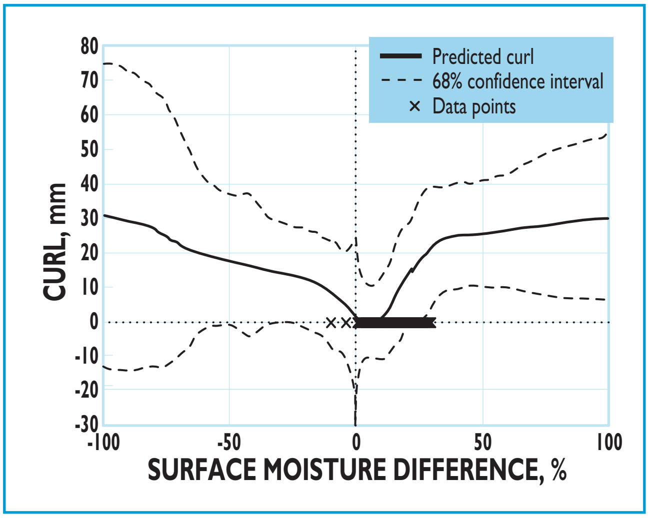
To assess the models and confidence measures, experiments were carried out by varying a single parameter while holding the others constant and by noting the prediction and confidence intervals. This was done for two parameters—surface moisture difference at the coating machine and caliper—noting that physically adjusting caliper is not feasible. However, here we are interested in the confidence intervals and how they react rather than in any physical interpretation. The original values of the parameters were 8.8% and 383 µm for the surface moisture difference and caliper, respectively.

The results of the experiment for the classification case are shown in **Figs. 3 and 4.** In addition to the prediction, the graphs also show 68% confidence intervals and training set data in that dimension. The confidence intervals relate to the certainty of the prediction—effectively stating that probabilistically, for the same inputs and 68% of the time, the prediction will be between these boundaries and will most likely be at the prediction level indicated. In other words, the wider the bounds, the less certain of the prediction we are.

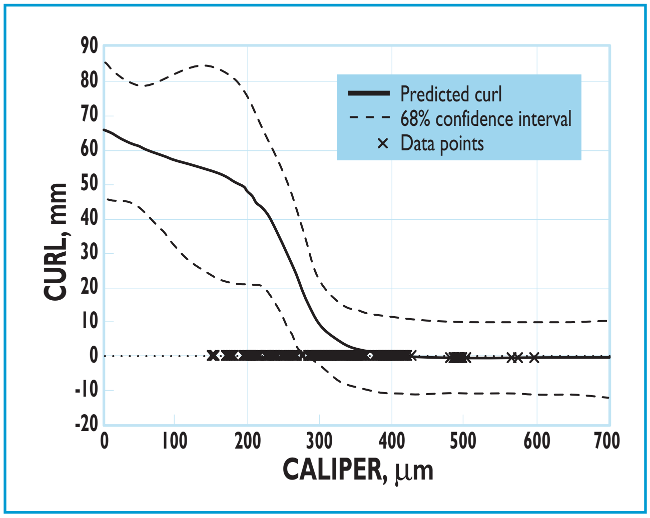
The graphs clearly show that as the parameters are adjusted, there is a change in the output prediction and, in addition, that as the parameters exceed (positively or negatively) the bounds of the training data, the confidence intervals widen. Clearly, as this is a high-dimensional problem, it is unclear exactly how the whole data set relates to the single dimension shown; however, the increasing width of the confidence intervals is encouraging. In addition, a physical interpretation of the results would suggest that the greater the caliper, the less likely it is that the paper will exhibit curl and that as the difference in the surface moisture increases in either direction, so does curl. While these interpretations also seem qualitatively reasonable, we can tell nothing of the predictive accuracy of the results—hence the need for confidence intervals.

A similar experiment was carried out for the regression networks (see **Figs. 5 and 6**), and generally encouraging results are seen. However, in the case of the caliper variation test (Fig. 6), the increase in the width of the confidence intervals is less marked because caliper varied in the positive direction. We believe that the reasons for this are related to the fact that outside the bounds of the database, the model is undefined. The confidence measure used to determine the 68% intervals used here is based on a method that calculates the uncertainty of the prediction due to noise in the data and to uncertainty in the model parameters. Generally, for this technique, the noise associated with the data is assumed to be constant, and any variation in the intervals occurs from variations in the parameter uncertainty.

For a part of the model that is poorly defined, i.e., when it is being used outside the bounds of the training data set, the prediction is



5. Variation in the prediction of curl as the surface moisture parameter changes. Also shown are the data points in that dimension and 68% confidence intervals.



6. Variation in the prediction of curl as the caliper variable changes. Also shown are the data points in that dimension and 68% confidence intervals.

uncertain, and this should be reflected in the confidence interval. However, the confidence interval in this case is also ill-defined and therefore inaccurate. While we have seen in these experiments that outside the bounds of the training data set, the intervals generally widen, here we see that they are also unreliable. This is an area of ongoing work. We note that where there is sufficient data, the confidence intervals as described are accurate and are an essential part of the curl prediction tool. However, further work is required to make the intervals reliable in all circumstances. Following the approach of Leonard *et al. (10)*, we are considering incorporating a component of the confidence measure that is directly related to the density of the data in the training set.

In summary, the work presented in this paper shows that ANNs can be used for predicting paper curl in the papermaking industry. In addition, we have shown that the perception that a neural network solution is a black box can be reduced via the use of confidence measures. Finally, we have developed a two-stage, decision-support tool that offers a prediction of the level of curl as well as a warning when a reel is likely to be out of specification.

CONCLUSIONS

Paper curl is an important quality measure in the papermaking industry and is typical of many similar metrics for the following reasons:

- The underlying mechanisms are not clear.
- The relationship between plant setting and quality metric is nonlinear.
- Historical data exist for training purposes, but data from the distant past is likely to be unrepresentative of the current conditions.
- Data are relatively sparse, "noisy," and often incomplete.
- Bad-quality paper wastes time and energy.
- Any decision-support tool must be simple to use and be trustworthy.

In this paper, we have shown that we can predict curl from parameters that define the current characteristics of a reel of paper and the plant machinery using neural network techniques. Specifically, we have developed models that can be used to predict (a) whether the characteristics of a paper reel will be out of specification with respect to curl and (b) the likely value of curl that will occur. In all the models, every prediction has an associated confidence interval to address the criticism that the neural network solution is a black box. These results are important in that they show how ANNs can be applied to complex and relevant practical problems in realistic working environments. **TJ**

Edwards is a research fellow, Murray is a professor, Papadopoulos is a research student, and Wallace is a lecturer at the University of Edinburgh, Electronics and Electrical Engineering Department, Edinburgh, EH9 3TL, Scotland, U.K. Barnard is coater development manager and Smith is a paper technologist at Tullis Russell & Co. Ltd., Markinch, Fife KY7 6PB, Scotland, U.K.

This work is supported by EPSRC grant no. GR/L30930 and by Tullis Russell & Co. Ltd.

Received for review June 8, 1998.

Accepted Oct. 30, 1998.

LITERATURE CITED

1. Bishop, C. M., *Neural Networks for Pattern Recognition,* Oxford University Press, New York, 1995, pp. 116–160.
2. Eriksson, L-E., Calvin, S., Fellers, C., *et al., Nordic Pulp Paper Res. J.* 2(2): 66(1987).
3. Goldner, P., *Tappi* 47(7): 168A(1964).
4. Langevin, E. T. and Giguere, W., *Tappi J.* 77(8): 105(1994).
5. Lyne, M. B., *International Printing and Graphic Arts Conference Proceedings,* TAPPI PRESS, Atlanta, 1988, p. 89.
6. Myles, A. J., "Methods for addressing some practical issues in MLP regression and their application to modeling curl in papermaking," Ph.D. thesis, University of Edinburgh, Edinburgh, 1997.
7. Tresp, V., Ahmad, S., and Neuneier, R., *Proceedings of the Neural Information Processing Systems (NIPS) Conference,* Morgan Kaufmann, San Francisco, 1994, p. 128.
8. Edwards, P. J., Murray, A. F., Papadopoulos, G., *et al., Proceedings of European Symposium on Artificial Neural Networks,* D-Facto, Brussels, 1999, p. 69.
9. MacKay, D. J. C., *Neural Computation* 4(3): 448(1992).
10. Leonard, J. A., Kramer, M. A., and Ungar, L. H., *IEEE Trans. Neural Networks* 3(4): 624(1992).