Advantages of using artificial neural networks techniques for agricultural data

N.B. Büchmann and I.A. Cowe

Foss Tecator AB, Box 70, Höganäs, Sweden

Introduction

The last decade has seen an increase in the use of artificial neural networks (ANN) as a mathematical tool in such diverse fields as medicine (identification of cancers), surveillance (identification of faces from video footage), communications (routing of telephone calls), finance (automated trading on the stock exchange) and biochemistry (identification of the tertiary structure of protein molecules). In this paper, we report some results for prediction of chemical composition of various agricultural products using ANN for both transmittance and reflectance near infrared data.

Use of ANN for ground samples of compound feed measured in transmission

Borggård[1] was the first to use ANN for agricultural data in the development of near infrared calibrations for controlling pig carcass quality. Encouraged by this work, Büchmann proposed the use of ANN as a modelling technique for near infrared transmission measurements of compound feed samples.[2] Three broadly based models are currently available for use with the Foss Infratec1275, one for use with compound feed for pigs and poultry, another for cattle and the third, typical feed ingredients.

The data sets used for these calibrations were extremely large; for example, the pig and poultry model was based on almost 18,000 records, which improved the robustness of the calibrations. Based on an independent prediction set of 2,270 samples, the following accuracy values (*SEP* corrected for bias) were achieved: crude protein: 0.59%; crude fat: 0.32%; moisture: 0.39%; and crude fibre: 0.51%. The accuracy for moisture was slightly disappointing, probably due to the fact that it is often difficult to take spectra immediately before drying, and changes may occur between taking spectra and reference analyses. The *SEP* values for the other parameters met commercially acceptable standards.

As the data for the ANN models were derived from more than ten instruments, instrument variation was, to a large extent, built into the model. Only a simple bias-adjustment is, therefore, required for calibration transfer to new units.

Use of ANN for crop data measured in transmission

The possibility of using ANN with transmission data from cereals for the determination of protein and moisture was first proposed by Büchmann in 1995.[3] Initial models were based on European data only. Later, Australian and North American data were added to the model and extensive testing has demonstrated the models can be applied worldwide. The European ANN models were first used commercially in 1996 by the Swedish Infratec Network. Since then, a number of European and Australian NIT networks have adopted the (now global) ANN models and extensive testing has been conducted in North America.

Near Infrared Spectroscopy: Proceedings of the 10th International Conference

In an evaluation with 481 US wheat samples, ranging from 8 to 22% protein, the *SEP* was 0.24%. With moisture, the improvement obtained by using ANN instead of PLS was very noticable. In a set of Danish barley samples, ranging from 13% to 26% moisture, the *SEP* for the PLS model was 0.75%, and predictions were non-linear over 18%. Using ANN, the *SEP* was reduced to 0.22% and the model was linear over the entire range. A comprehensive comparison between the PLS and ANN models for Danish crops has been published elsewhere.[4]

A single global ANN model for protein is suitable for wheat, barley and rice, while the moisture model can be used with wheat, barley, rye, oats and triticale. Currently, ANN models are in development for moisture and oil in rapeseed, for moisture and amylose in rice and for moisture and oil in maize. These models are, as yet, still restricted in terms of geographical coverage and cannot be considered completely global at this stage. The data set for cereals now consists of nearly 40,000 records including samples from the last 15 years, obtained from all over the world and measured on Foss Infratec types 1221, 1225, 1229, 1241 and 1275. With such extensive sample variation built into the database, an annual update of the global models is sufficient. After updating, the performance of each new model is then evaluated with independent data sets from each NIT network and network-specific bias adjustments are applied where appropriate. A total integration of ANN models across different networks is, thus, still one step away from completion. This, however, is not due to any shortcomings of the global ANN models themselves; imperfect harmonisation of the reference analyses used at the individual control laboratories is certainly a major influence.

Characteristic differences between Infratec transmission and NIR reflectance spectra

The applications mentioned above all relate to Foss Infratec instruments, where the spectra consist of 100 data points over a narrow wavelength range from 850 nm to 1050 nm measured in transmission. Reflectance spectra from the Foss NIRSystems (NIR) 5000 series consist of 700 data points over a range from 1100 nm to 2500 nm. NIR spectra are thus more complex than those from Infratec instruments and, because of their greater range, tend to display more non-linearity. However, absorbance levels measured in transmission are, typically, in the range between 2 to 4 O.D., where similar samples measured in reflectance rarely exceed 1.2 O.D. Transmission data, therefore, tend to show greater deviation from Lambert–Beer's Law compared with reflectance spectra.

Use of ANN for ground samples of compound feed measured in reflectance

Most of the major Infratec transmission applications are based on models developed centrally, using ANN combined with a Foss Tecator proprietary pre-processing technique. ANN as a calibration tool has never been extensively used for agricultural data measured in reflectance, perhaps because non-Foss expert users developed these applications and, without the proprietary pre-processing technique, results may have been disappointing. Over the last year, however, Foss Tecator and ISI initiated extensive evaluations of modified partial least squares (MPLS) regression, a proprietary ISI technique, and ANN as alternative calibration methodologies for reflectance data from FossNIRSystems instruments. An evaluation of the proprietary ISI technique, LOCAL, was an equal part of the studies, but these results are outside the scope of the current proceedings. One study focused on dry, ground forage data from Australia, Europe and North America. Results relating to this study will be presented separately by Berzaghi[5] and Dardenne[6] at this conference. The second study was carried out on compound feed and feed ingredients of plant origin using historical data from Australia, Europe and North America. Some of the findings with compound feed and feed ingredients will be discussed here.

Historical data derived from Foss NIRSystems instruments trimmed to NIRS5000 format (1100 to 2500 nm) were obtained from Australia, Belgium, Holland, North America and the UK. All of these instruments were standardised to the ISI Master (Ser. No. 1272). The combined modelling set consisted of 6,115 records. All types of feed and feed ingredients were combined into a single ANN model. An independent data set for validation of the models consisting of 530 records of mixed feed collected on an instrument that was not standardised against the ISI Master was obtained from a single Italian site. The *SEP* values for the independent data set were 1.12% for crude protein and 0.79% for crude fibre.

Although these values are not accurate enough to be commercially acceptable, the study does indicate that data from very diverse samples can be combined in broadly-based models. Several factors should be taken into account. Firstly, the independent test sets were not standardised to the ISI master, and hence not to the data set used to derive the model. Secondly, even though the spectra used to derive models were standardised, these standardisation exercises were not always performed using to the same protocol. Finally, there is no certain way to ensure that the different, historical sources of reference data are properly harmonised. The study did, however, confirm that correcting for current laboratory differences by means of a ring test did improve the performance of the resulting models.

Use of ANN for unground samples of compound feed measured in reflectance

Until recently, near infrared measurements of compound feed were typically carried out on samples ground to a 1 mm particle size (reflectance) or 2 mm size (transmission). However, near infrared measurements of such products can be further simplified by omitting grinding altogether and measuring whole feed pellets directly. As part of the calibration methodology study with feed, we compared PLS with ANN on a set of unground samples of feed for pigs, cattle and poultry (both broiler and layer). A separate MPLS model was developed for each of these feed classes, while the ANN model was based on all four classes combined. The ANN calibration set consisted of 6,142 protein, 4,860 fat and 4,967 moisture values. The results, expressed as the *SEP* values averaged over the different feed classes, are shown in Table 1. These results indicate that ANN, on average, offered a 10% improvement over PLS. In addition, the combined ANN model is easier to use and support compared with several individual PLS models. While not as low as the corresponding *SEP* figures for ground samples (data not shown here), the *SEP* values for whole pellets are sufficiently good for practical use (Jos Zegers, personal communication).

Use of ANN for molasses measured in reflectance

ANN has also been studied briefly for use with liquid samples. Purity of molasses is defined as Pol*100 Brix^{-1}, where Pol is the optical rotation and Brix the soluble solids. ANN and PLS calibration models developed from approximately 1,000 samples were evaluated using an independent dataset with 80 records. With PLS, the *SEP* for purity was 0.40, too high for commercial use; with ANN, the *SEP* improved to 0.30, approaching the limit for a useful calibration (Monique Seegmans, personal communication). However, further testing with the ANN model indicated that the *SEP* obtained was not reliable, probably because a calibration set of only 1,000 records was lower than the size usually considered necessary for ANN to produce stable models.

Table 1. Accuracy of ANN vs PLS models for unground samples of compound feed.

% improvement in *SEP* of ANN over PLS				
	N	ANN	PLS	% improvement
Protein	710	0.89	1.05	16
Fat	580	0.60	0.62	3
Moisture	546	0.35	0.39	11

Spectral robustness of ANN and PLS models

David Funk has studied the effect of deliberate spectral pertubations on the prediction output of various calibration models for wheat protein (personal communication, extracted from an unpublished presentation given at a GIPSA NIRT Calibration/Standardisation Meeting, May 16–17 2000, Kansas City, Kansas, USA).

The perturbations included absorbance offset, pathlength increase, stray light addition, bandwidth changes, wavelength axis stretch, wavelength axis offset and wavelength stretch offset. The models assessed were six separate official GIPSA PLS calibrations compared with the global ANN model from Foss Tecator (WHPR2121). Funk concluded that the ANN model was less sensitive to most disturbances than the GIPSA PLS calibrations. Thus, the ANN model was more able to cope with instrument differences than the PLS models. In a broader context, the study is an elegent (although indirect) approach allowing a comparison between the behaviour of non-linear and linear calibration methods.

Conclusion

Some distinct features of ANN techniques for agricultural data can be summarised as follows:

- ANN, because it is a non-linear technique, is better than PLS where the range of concentration is large where there is non-linearity in the dataset.
- ANN can combine diverse samples into broadly-based models, simplifying calibration support.
- Although ANN is often vastly superior to PLS when dealing with high absorbance levels in transmission measurements, ANN has also consistently been found to be more accurate than PLS with reflectance data.
- ANN requires very large data sets; real improvements can still be found with more than 20,000 spectra. While such datasets are costly to procure and complicated to compile, they still offer the best solution to the problems of long-term stability and cost-efficient calibration support.
- Enhancement of ANN models only required a few additional samples and existing levels of performance were not upset by the new data.
- Harmonisation of laboratories improved performance of the models, regardless of whether they were based on MPLS or ANN. However, historical datasets can not be reliably harmonised because the records required rarely exist.
- ANN is less sensitive to most spectral pertubations than PLS and can cope better with instrument differences.

Acknowledgements

The molasses results are published with the generous permission of Mrs Monique Steegmans, Tiense Suikerraffinaderij-Orafti, Tienen, Belgium while those for unground compound feed pellets are published with the generous permission of Mr Jos Zegers, Maasweide Laboratory Services, Boxmeer, The Netherlands. The results for spectral robustness are extracted with the generous permission by Mr David Funk, GIPSA (USDA), Kansas City, MO, USA.

References

1. C. Borggaard and A.J. Rasmussen in *Near Infrared Spectroscopy—Bridging the gap between data analysis and NIR applications*, Proc. of the 5th Int. Conf. on NIR Spectrosc., Haugesund, Norway, Ed by K.I. Hildrum, T. Isaksson and T. Næs. Ellis Horwood, Chichester, UK, p. 73 (1992).

2. N.B. Büchmann, in *Leaping ahead with Near Infrared Spectroscopy*, Ed by G.D. Batten, P.C. Flinn. L.A. Welsh and A.B. Blakeney. Royal Australian Chemical Australian Institute, Victoria, Australia, p. 248 (1996).
3. N.B. Büchmann, in: *Near Infrared Spectroscopy: The Future Waves*, Ed by A.M.C. Davies and P. Williams. NIR Publications, Chichester, UK, p. 479 (1996).
4. N.B. Büchmann, H. Josefsson and I.A. Cowe, *Cereal Chemistry* **78(5),** 572 (2002).
5. P. Berzaghi, P.C. Flinn, P. Dardenne, M. Lagerholm, J.S. Shenk, M.O. Westerhaus and I.A. Cowe, *Changing the world with NIR, (*in press).
6. P. Dardenne, I.A. Cowe, P. Berzaghi, P.C. Flinn, M. Lagerholm, J.S. Shenk and M.O. Westerhaus in *Changing the world with NIR*, (in press).