

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/263502603>

Modular Neural Network Modelling for Long-range Prediction of an Evaporator

Article in *Control Engineering Practice* · January 2000

DOI: 10.1016/S0967-0661(99)00123-9

CITATIONS

15

READS

24

3 authors, including:



[Huub H. C. Bakker](#)

Massey University

26 PUBLICATIONS 151 CITATIONS

[SEE PROFILE](#)

Paper for publication in Control Engineering Practice

Paper Number: CD 807

Title: Modular Neural Network Modelling for Long-Range Prediction of an Evaporator

Running title: Modular Neural Network Modelling for an Evaporator

Authors: N.T. Russell[†], H.H.C. Bakker[‡] & R.I. Chaplin[‡]

Authors' Affiliation: [‡]Institute of Technology & Engineering, Massey University,
Palmerston North, New Zealand

[†] Predictive Control Ltd, Highbank House, Exchange St, Stockport,
Cheshire SK3 0ET, UK
Fax: +44-1606-44592

Contact Author: Dr. H. Bakker
Institute of Technology & Engineering
Massey University
Private Bag 11222
Palmerston North
NEW ZEALAND
Fax: +64-6-3505604
Email: H.H.Bakker@massey.ac.nz

Abstract

This paper presents the development of a modular neural network model of a three-effect, falling-film evaporator. The model comprises a number of sub-networks each modelling a specific element of the overall system. The modular structure was employed in order to provide benefits in terms of improved model training and performance. The performance of the modular neural model is demonstrated for long-range prediction by comparing it with process data, an analytical simulation and a linear ARX model. The results show that the modular neural model can satisfactorily predict over a horizon of arbitrary length and is suited for implementation within a predictive control scheme. Benefits in terms of model flexibility and interpretability are also discussed.

Keywords: Neural networks; simulation; prediction; modular modelling; evaporators; model-based predictive control.

Modular neural network modelling for long-range prediction of an evaporator

N.T. Russell[†], H.H.C. Bakker[‡] & R.I. Chaplin[‡]

[‡] *Institute of Technology & Engineering, Massey University, Private Bag 11222, Palmerston North, New Zealand*

[†] *Predictive Control Ltd, Highbank House, Exchange St, Stockport, Cheshire SK3 0ET, UK*

Abstract

This paper presents the development of a modular neural network model of a three-effect, falling-film evaporator. The model comprises a number of sub-networks each modelling a specific element of the overall system. The modular structure was employed in order to provide benefits in terms of improved model training and performance. The performance of the modular neural model is demonstrated for long-range prediction by comparing it with process data, an analytical simulation and a linear ARX model. The results show that the modular neural model can satisfactorily predict over a horizon of arbitrary length and is suited for implementation within a predictive control scheme. Benefits in terms of model flexibility and interpretability are also discussed.

Keywords: Neural networks; simulation; prediction; modular modelling; evaporators; model-based predictive control.

1. Introduction

Artificial neural networks (NNs) are a parallel processing paradigm inspired by the way biological nervous systems learn and process information. They have become useful tools for process identification because of their ability to represent nonlinear systems and their easy development (relative to conventional methods) using measured process data.

With the advent of artificial NNs the cause of nonlinear process control has been significantly advanced. Neural modelling techniques have been implemented within a range of control strategies including inferential control (Willis et al., 1991), model reference adaptive control

* Author to whom correspondence should be addressed – *facsimile: +64-6-3505604; email: H.H.Bakker@massey.ac.nz*

(Narendra & Parthasarathy, 1990), internal model control and model-based predictive control (Hunt et al., 1992; Morris et al., 1994).

In recent chemical process control research model-based predictive control (MBPC) schemes have become one of the most common control methodologies to which NNs have been applied. In general, model-based predictive control is a receding-horizon strategy which aims to optimise the control moves over a future horizon based on a desired control objective. The controller makes use of an explicit dynamic process model to estimate the process outputs over the prediction horizon. The model plays a vital role in the operation of the controller and, in effect, determining an appropriate model to use constitutes the majority of the controller design.

A model used within MBPC is required to be a dynamic model with the ability to predict ahead in time, referred to as n -step-ahead prediction, where n is the length of the finite prediction horizon. Some dynamic models perform well in estimating over a short horizon but when faced with a more realistic and useful horizon of say 20 or more time steps their performance can degrade rapidly. This is particularly true for NN models. The ability to predict ahead must therefore be built into the training methodology. Often the exact length of prediction horizon is not known prior to the training of the NN model therefore it is sensible to train the network to predict over the entire range of the available data. A data set which spans N time steps would then be used to train a network to predict up to N steps ahead. For the purposes of this paper this is referred to as *long-range* prediction. Using the network to produce long-range predictions is equivalent to implementing the network as a pure simulation model where only past information up to time t is used to predict up to time $t + N$.

In general NNs are time-independent models. To capture the dynamics of a system time-delay networks are commonly used (Narendra & Parthasarathy, 1990).

These networks can however dramatically increase the dimensions of the NN, which can decrease the speed and performance of the training method. The use of modular modelling approaches can alleviate these problems. Additionally, by combining prior knowledge of the

system to be modelled within a modular structure, one can increase the transparency and interpretability of the trained NN model.

2. The evaporation process

The evaporation process is a good example of a process that requires accurate control. A large proportion of energy used in industry is given to drying processes and evaporators play a significant role in the industrial drying of a number of food products like milk powders. Good evaporator control is particularly important since evaporators are often located directly upstream from energy-intensive processes such as spray drying. Tight control on the evaporator leads directly to better control in the dryer which results in better energy efficiency and a more consistent product.

The subject of this study is a pilot-scale evaporator resident within the Institute of Technology & Engineering, Massey University. The evaporator is a three-effect, falling-film evaporator with two preheater stages and a condenser. A schematic of a single evaporation stage (effect) of the pilot-plant is shown in Figure 1 **Fig. 1**.

Industrial-scale evaporators have many evaporation tubes per effect, however the pilot-plant has just a single tube. The product flow enters the top of the effect through a distribution nozzle and plate arrangement and flows down the evaporation tube as a film, boiling off as it descends. The vapour is separated from the liquid and is drawn off to be used as the heating medium for the downstream effect. The concentrated product forms a level in the reservoir at the base of the effect and is feed to the next effect for further concentration. The pilot-plant evaporator has yet to be operated with liquids containing a solute, using instead pure water as the product stream. The product concentration is not therefore considered in this paper. Incorporating equations for the concentration of a solute would increase the nonlinear characteristic of the model.

The variables of interest are the product temperature in the effect, the concentrate level in the reservoir and the product flowrate out of the effect.

An analytical model of the pilot-plant evaporator was developed as a simulation tool (Russell, 1997). This model was used to compare with the empirical models described later in this paper. The approach used for the analytical evaporator model development was taken from the work of Quaak and Gerritsen (1990). They developed a dynamic model for a multi-effect evaporator and used a systems approach in which the process was divided into sub-systems for the purpose of analysis. The sub-systems included the model distribution plate, evaporation tube, product transport, and energy flows for each effect. The states of the model are the concentrated product temperatures, flowrates and levels. The model of the Massey University evaporator extended Quaak and Gerritsen's model to include additional sub-systems relating to feed, preheater and condenser systems.

The settling times for the pilot-plant were approximately 1 minute for the concentrate levels and of the order of 2 to 4 minutes for the effect temperatures. It was assumed that the time constants for the flowrates were negligible. Data from the plant was sampled at 5 second intervals.

3. Neural network model development

3.1 Modular modelling approach

The objectives in the structuring of the NN model were to

- simplify the training task as much as possible, and
- develop a model with a more meaningful structure.

To this end it was decided to create a modular model which made use of prior knowledge of the system. More specifically the NN model uses prior knowledge to create a modular model in two ways:

- i) The modelling task was decomposed into smaller elements or modules representing sub-units within the total system, similar to that used in the analytical model. The sub-networks were combined to form the full model according to the structure of the actual process. Each network was used to predict only one output variable. The inputs were

Modular neural network modelling for long-range prediction of an evaporator

chosen to be those variables that were known to strongly influence the output and could include outputs of other sub-networks.

ii) Following the approach of Mavrovouniotis & Chang (1992) each sub-network was further simplified through localised computation by grouping related inputs together. The structure of the sub-networks were determined using the following guidelines:

- The delayed inputs for each variable were connected to two hidden neurons.
- All the inputs relating to a particular time instant were also connected to a single hidden neuron.
- All the hidden neurons were connected to a single output neuron.

These guidelines represent a simple and compact approach to sub-network structuring. The two hidden neurons coupled to each input attempt to capture the dynamics inherent within each input variable. Using only one neuron for this task would not be sufficient to describe complex behaviour whereas using more than two neurons is unlikely to significantly improve the characterisation.

The additional hidden neurons attempt to capture the inter-relationship between the inputs at each time instance. These time-related hidden neurons also prevent the network structure becoming simply a parallel combination of the inputs and can be considered to provide a 'snap-shot' of the system inputs at a particular time.

An example of the structure of a sub-network with localised connections and output feedback is illustrated in Figure 2. The output feedback is necessary to produce n -step-ahead predictions since past outputs need to be available for the model at future times steps. This structure of network is known as an *externally-recurrent* network (Su *et al*, 1992).

The benefits of a modular modelling approach using prior knowledge are three-fold:

- i) to provide an efficient method for tackling large modelling problems, resulting in improved training and generalisation;

- ii) to provide flexibility so that the model can be easily updated and modified as the need arises and allows differing model characteristics and methodologies to be included within the single overall model;
- iii) to give structure and an element of transparency to a NN so that the model can be more easily analysed and understood with respect to the actual system.

Point iii) is especially appealing since NNs generally have an amorphous structure whose relevance to the physical system they represent is not apparent. It is also difficult to analyse the dependencies occurring within in a network and how the outputs are calculated.

3.2 *Sub-network training methodology*

The network training method developed for this study combined a backpropagation through time procedure (BPTT) with the Levenberg-Marquardt optimisation technique (Marquardt, 1963; Levenberg, 1944). The name Levenberg-Marquardt through time (LMTT) was chosen to describe this combined methodology.

The backpropagation through time (BPTT) approach (Rumelhart *et al*, 1986; Werbos, 1990) uses data consisting of N time steps to train an externally-recurrent network by unfolding the network N times to create a feedforward network where the weights in each layer are identical. A modified version of the standard backpropagation (BP) algorithm is used to update the weights of the network. Previous work of a similar nature to this current study has employed conjugate gradient and random search techniques with BPTT to overcome poor convergence (Su *et al*, 1992). The Levenberg-Marquardt optimisation method (LM) offers a more efficient solution for network training. LM can easily be incorporated into the backpropagation method replacing steepest descent minimisation and its superior performance over steepest descent and conjugate gradient learning in NNs has been demonstrated (Hagan & Menhaj, 1994). However, its main disadvantage is that significantly more processing memory is required compared with other methods.

When using such a recurrent training approach it is necessary during the presentation phase that the input matrix is updated with the network predictions. The training data is presented in

batch mode and the network predictions are used to update the input vectors. This ensures that the network is trained using a parallel method (Narendra & Parthasarathy, 1990) and the resulting errors are calculated based on a recurrent implementation.

There are alternative approaches to BPTT for training recurrent networks (Pearlmutter, 1990) which have been developed for fully recurrent networks. These alternative methods are often more efficient than BPTT since they are able to train networks without the need for unfolding the network. These approaches have a higher computational cost than BPTT but are more memory efficient. A BPTT approach was chosen because of its simplicity and easy adaptation from the standard BP algorithm.

4. The evaporator neural network model

4.1 Process data

Pseudo-random sequences were applied to the inputs of the evaporator simulation in order to excite the model states and demonstrate the dynamics of the evaporator. For nonlinear identification the response data needs to represent the full range of the system dynamics and therefore a larger number of operating levels were applied to each input variable. Three separate trials (carried out on different days) were run in order to obtain three independent data sets. These three data sets were then assigned to be the training, testing and validation sets for the evaporator modelling task.

The random inputs were simultaneously applied to the feed pump, each of the effect pumps and the steam valve to produce dynamically rich response data. A sampling rate of 10 seconds for the training data was considered appropriate for the short time constants present in the evaporator process. The training data set was selected as the set with the widest variation in process outputs. This data set covered an interval of over two hours. The testing set was 90 minutes long and was also used during the training phase. The validation data covered an interval of 53 minutes and was used to give an unbiased estimate of the model performance.

4.2 Structure selection

The network structures were chosen on the basis of known relationships through experience with the process and the analytical model development. Cross-correlation tests were also performed both to assist in selecting the correct cause-to-effect relationships and to determine the delay spread of the time-delayed inputs into the networks.

Similar network structures have been used in previous work related to the pilot-plant evaporator which made use of test data from the analytical simulation (Russell & Bakker, 1997). The model structures used were found to be successful and the study proved a useful preliminary study.

The choice of the number of hidden neurons was not necessary for the most of the sub-networks since their structures were determined by the guidelines described in Section 3.1. Therefore the number of hidden neurons is dependent on the number of input neurons. The only exceptions to this were the flowrate sub-networks which were only first order systems, had no time delayed input streams and therefore did not follow the structure rules outlined in Section 3.1.

Some examples of the second effect sub-networks will help to demonstrate the structures used. Each network was used to predict only one output variable and its inputs are variables that are known to strongly influence that particular output. The resulting sub-nets have multi-input, single-output structures.

The following equations define the functionality of the three sub-networks that make up the effect model.

- The effect two temperature was selected to be a function of the upstream and downstream effect temperatures (T_1 and T_3), the flowrate into the effect (Q_1) and previous values of the effect two temperature.

$$T_2 = f(T_1, T_3, Q_1, T_2) \quad (1)$$

These variables were selected since the upstream and downstream temperatures directly influence the effect temperature and the mass of product in the evaporation tube strongly

Modular neural network modelling for long-range prediction of an evaporator

influences the rate at which the temperature changes. After investigating cross-correlation results and testing a number of model structures it was decided that each input to the network would have the same delay spread of four samples. The number of hidden neurons used was determined by the guidelines in Section 3.1.

- The effect two concentrate level was chosen to be a function of the flow in (Q_1) and out (Q_2) of the effect, the temperature of the effect (T_2) and previous values of the level.

$$L_2 = f(Q_1, Q_2, T_2, L_2) \quad (2)$$

These variables were selected since, according to the analytical model, the change in the level is a function of the product flow into the effect and the product and vapour flows out of the effect. The temperature has also been included as a variable to represent the vapour flow since the higher the temperature the greater the evaporation. After investigating cross-correlation results and testing a number of model structures it was decided that each input to the network would have the same delay spread of three samples. The number of hidden neurons used was determined by the guidelines in Section 3.1.

- The flow out of effect two was determined as a function of the differential pressure between effect two and three which is estimated using the temperature difference $T_2 - T_3$, the pump speed (N_2), the concentrate level (L_2) and the previous value of the flowrate.

$$Q_2 = f(T_2 - T_3, N_2, L_2, Q_2) \quad (3)$$

These variables were selected after considering the energy balance across the product transport system. Using an input for the pressure difference in this network, rather than the temperature difference, did not produce an improved performance. Since the flowrate dynamics are fast it was decided that each input to the network would have a delay spread of 1 sample – essentially producing a static model. This meant that the flowrate network topology would not be determined according to the rules of Section 3.1. Instead the number of hidden neurons used was determined by performing selection trials where the

Modular neural network modelling for long-range prediction of an evaporator

number of neurons was varied between two and twelve with at least ten networks tested at each number of hidden neurons. The structure that minimised the testing error was selected. For the four flowrate sub-nets in the evaporator model either five or six hidden neurons were used.

Each sub-network had sigmoidal hidden-layer neurons and linear output neurons. Sigmoidal type networks were used since similar recurrent network training had been carried out using this type of network (Su et al., 1992).

4.3 Network training

All training and testing data was scaled to lie within the range -1 to $+1$.

The network training was applied individually to each sub-network. As the network weights were updated the testing data was passed through the network to provide a measure of the network's generalisation capabilities during the training process. The network with the lowest testing prediction error was selected as the 'best' model. The test for each network was to perform a long-range prediction over the whole testing data set. This equates to 90 minutes of data.

4.4 Complete evaporator network model

The complete evaporator neural model consists of thirteen sub-networks each describing a separate state of the three-effect system. The model states include the product temperatures, concentrate levels and product flowrates throughout the evaporator. The inputs are the steam valve position and the speeds for the feed pump and each of the effect pumps.

Once trained the sub-networks were combined to form the complete evaporator model. The general structure of the complete evaporator model is illustrated in Figure 3. The complete model consists of the thirteen sub-networks connected in parallel and is equivalent to a sparsely-connected two-layered network with an additional input layer. Any secondary relationships between the system variables that are not modelled within the sub-networks are

still accounted for in the manner in which they are joined together since the outputs of some sub-networks are used as inputs for others.

5. Alternative models

Additional models of the evaporator were developed which employed alternative structures or modelling methods in order to provide a comparison for the modular NN model.

5.1 Other neural network models

NNs with alternative topologies were developed to compare with the locally-connected, modular sub-networks to determine whether the proposed modular structures produced improved results. The alternative networks included:

- Fully connected sub-networks to compare with the locally-connected sub-networks for a sub-section of the evaporator.
- Locally-connected, three-output NN to compare with the modular-structured NN model for a sub-section of the evaporator.

The second effect of the evaporator was selected as the sub-section to be modelled for the above comparisons. The alternative models were fed the identical inputs used by the corresponding modular sub-networks.

The number of hidden nodes used in the fully-connected sub-networks were chosen so as to give a similar number of weights as the locally-connected sub-networks (Table 1).

The number of weights used in the three-output network with a locally connected structure was less than that of the modular structured network as shown in Table 2.

5.2 Linear regression model

A second modular model was developed for the full evaporator system with an equivalent overall structure to that of the modular NN model of Figure 3 but with the nonlinear sub-networks being replaced with linear ARX (AutoRegressive with eXogenous inputs) regression sub-models. These sub-models were identified using the least squares approach of Ljung (1987).

6. Results and discussion

All the models were cross-validated using an independent data set from the actual process to perform an unbiased test. All the following results are based on comparing the model performances to this validation data set.

6.1 Comparison of locally-connected sub-networks and alternative networks

The locally-connected sub-networks exhibited improved training and superior long-range prediction performance over the sub-networks with a fully-connected topology. The training of the locally connected sub-networks generally converged to more reliable solutions and more rapidly than the training of the alternative models (Russell, 1997). The scaled mean-square errors (MSEs) for the second effect product temperature and concentrate level predictions are shown in Figure 4. As well as the improved performance the locally-connected networks have fewer network parameters and allow the network behaviour to be more readily analysed and understood. Results for the flowrate sub-network are not shown since it is a first order model and therefore does not have a locally-connected structure.

Similar results were obtained when comparing the training and performance of a single three-output NN and the modular model consisting of three sub-networks for the second effect (Russell, 1997). The MSE for the long-range predictions are plotted in Figure 5. Despite having fewer network parameters the three-output NN model does not generalise as well as the modular network model.

6.2 Performance of complete evaporator models

Implementing the modular NN and ARX models to produce long-range predictions allows a direct comparison to be made with a nonlinear analytical simulation of the evaporator (Russell, 1997).

The MSE results for the long-range prediction of the validation data for each variable of interest are plotted in Figure 6. Generally the performance of all the models were similar in character in that the temperature (T_i) and flowrate (Q_i) estimates were closer to the plant data

while the concentrate levels (L_i) were the hardest to model accurately. It is clear from the results that both the NN and ARX models out-perform the analytical model in estimating the plant data over the majority of the variables. It is also interesting to note that the linear ARX model error results are generally similar or better than that of the nonlinear NN. This leads one to conclude that the data may not, in fact, be highly nonlinear (since the process was not operated with a product containing solids) and that some of the sub-networks did not capture all the nonlinear behaviour. However one must investigate plots of the model responses before making such conclusions since the MSE results only give a partial indication of the model fit.

By way of example two plots of the model predictions are illustrated below.

Figure 7 shows a plot of the actual second effect temperature (T_2) compared with the predicted outputs of the NN and analytical models. Both models perform well in matching the actual data. The ARX model has a similar response to the two that are plotted.

All the models had difficulty in predicting the second effect concentrate level (L_2) accurately over the entire time range. The empirical models however out-perform the analytical model. While the NN model has a worse MSE result than the ARX model it is mainly due to its poor response at the low values of the level during the period of 30 to 40 minutes (Figure 8). The remainder of the NN model estimate is superior to that of the ARX model having a closer fit. Subsequent analysis of the training data for the L_2 sub-network showed that it did not span an adequate region of the variable range. This highlights one problem when training NN models to predict a number of variables; it can be difficult to obtain process data containing adequate excitation of all the variables of interest.

Figure 7 is an example of a typical prediction result whereas L_2 (Figure 8) is the poorest predictor from the complete model.

6.3 Network performance versus prediction horizon

To test whether the modular neural model could predict over an arbitrary length horizon the model was simulated with varying lengths of prediction horizons in order to predict the

validation data. Figure 9 shows a plot of MSE prediction error versus the horizon length, n , for the temperature variables. The trends show in general that as the prediction horizon is increased the errors initially increase but then reach a plateau where the error remains approximately constant for increased horizons. The exception to this appears to be the second effect temperature. However the plots appear to indicate that the network does allow an arbitrary length prediction horizon to be used without significant degradation of the model predictions. Plots exhibiting similar trends were also obtained for the flowrate and level predictions.

6.4 Model flexibility

The flexible nature of the model structure allows the development of a mixed-model representation of a system where various alternative modelling techniques can be applied to represent different portions of the system. Some sub-systems may be modelled with nonlinear NNs while others may only require a linear model. The sub-networks could also be incorporated with analytically derived models to enhance the model where theoretical process knowledge is lacking.

A modular structure also allows modifications and extensions to the model to be easily carried out if changes in the process need to be taken into account, thus reducing future model development effort. If a particular sub-network is performing poorly than this element can be singled-out for re-development.

6.5 Model transparency

As an example the T_2 locally-connected sub-network can be used to illustrate how a measure of interpretation can be given to the relationships between the inputs and the T_2 output. The outputs of the three time-based hidden-layer neurons are plotted in Figure 10. The inputs of one neuron are all the input variables at time $t - 1$, another connects with the variables at time $t - 2$ and the other at time $t - 4$. From this plot it appears that different neurons are

characterising different variations in T_2 . The inputs at $t - 2$ produce a fairly constant output while the inputs $t - 4$ model the major variations in the output, T_2 .

While it is not obvious from Figure 10 what the precise relationships are, the structure allows one to more easily break down the relationships within the overall network and discriminate between the effects of the different inputs.

6.6 Training effort

The locally-connected sub-networks trained better than fully-connected networks, which were prone to becoming trapped in local minima causing the LMTT method to terminate prematurely.

Training times for the locally-connected sub-networks were generally less than 100 epochs and converged to a reasonable solution whereas the fully-connected sub-networks typically took greater than 400 epochs and still failed to converge.

Similar problems were experienced with the three-output network despite having a localised structure. However the training and generalisation capabilities of this network were superior to a fully-connected, three-output network which was also developed.

7. Conclusions

A framework for developing a dynamic NN model, made up of several sub-networks, has been presented which enables prior process knowledge to be built into the model topology. This approach of breaking the model into sub-systems removes the need to optimise a problem of high dimension. The modular structure of the model also provides benefits in terms of model flexibility and transparency.

It was found that generally the more complex the structure the more difficult it was to train effectively and hence networks with poorer predictions were produced. The suggestion that ‘the more network weights—the harder to train’ holds true when comparing the training and performance of the fully-connected and locally-connected sub-networks. However the modular model out-performed the three-output network despite it having more weights. This is due to the fact that the individual sub-networks were easier to train individually and when linked together performed better than the single network structure.

Based on observations and results from this research it can be stated that using:

- localised connections reduces the number of network parameters which benefits training and provides an element of improved model interpretability, while
- modularisation improves network prediction performance and offers flexibility.

The complete evaporator NN model performed long-range predictions satisfactorily when presented with a validation data set and had a superior performance to that of a nonlinear analytical simulation of the pilot-plant. Compared with an equivalent linear ARX model the NN model had comparable or slightly worse error prediction results. For a nonlinear chemical process one would expect that the nonlinear NN would perform better. These results lead to the conclusion that either:

- the process is not highly nonlinear and linear models are sufficient for this application, or
- some of the sub-networks failed to converge to a satisfactory solution

The absence of highly nonlinear data could be due to the fact that the evaporator was operated with pure water rather than a product like milk, which has complex physical properties.

The second point raises the question of how to improve the convergence for recurrent training. Some improvement has been experienced when a teacher forcing technique is applied to the training (Williams & Zipser, 1989; Pearlmutter, 1990). With this scheme the target output is used to drive the network dynamics in place of the output feedback. This can help to improve network convergence but can also cause instability in the responses.

Another solution which may improve the accuracy of the modular NN is to apply a global optimisation to the model after individually training the sub-networks (Gallinari, 1995). This could help reduce the accumulation of errors through the model but would require a more efficient training algorithm than used here.

Either way, more work is required to investigate these possibilities further. Both in terms of using data from trials using a more realistic product stream and investigating alternative modular network-training strategies. These will be the focus of further research.

Despite this, the prediction results shown here do demonstrate that the modular NN model can model a chemical process and generate satisfactory predictions over an arbitrary prediction horizon. This type of model therefore has the potential to be implemented within a model-based predictive control strategy.

8. References

- Gallinari, P. (1995). Training of modular neural net systems. *The Handbook of Brain Theory and Neural Networks*, (M. A. Arbib, Ed.), MIT Press, MA, USA, pp. 582-585.
- Hagan, M.T. & Menhaj, M.B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Trans. on Neural Networks*, **5**, 989-993.
- Hunt, K.J., Sbarbaro, D., Zbikowski, R. & Gawthrop, P.J. (1992). Neural networks for control systems—a survey. *Automatica*, **28**, 1083-1112.

Modular neural network modelling for long-range prediction of an evaporator

- Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least squares. *Quarterly of Applied Mathematics*, **2**, 164-168.
- Ljung, L. (1987). *System Identification - Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, USA.
- Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal. Soc. Indust. Appl. Math.*, **11**, 431-441.
- Mavrovouniotis, M.L. & Chang, S. (1992). Hierarchical neural networks. *Computers & Chem. Engng.*, **16**, 347-369.
- Morris, A.J., Montague, G.A. & Willis, M.J. (1994). Artificial neural networks: Studies in process modelling and control. *Trans. Inst. Chem. Eng.*, **72**, 3-19.
- Narendra, K.S. & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Trans. on Neural Networks*, **1**, 4-27.
- Pearlmutter, B.A. (1990). Dynamic recurrent neural networks. Technical Report CMU-CS-90-196, Carnegie Mellon University, Pittsburgh, PA.
- Quaak, P. & Gerritsen, J.B.M. (1990). Modelling dynamic behaviour of multiple-effect falling-film evaporators, *Computer Applications in Chemical Engineering*, (H.T. Bussemaker & P.D. Iedema, Eds.), Elsevier, Amsterdam, pp. 59-64.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation, *Parallel Distributed Processing*, (D.E. Rumelhart, & J.L. McClelland Eds.), MIT Press, Cambridge, MA, USA, Vol. 1, pp. 318-362.
- Russell, N.T. (1997). Dynamic Modelling of a Falling-Film Evaporator for Model Predictive Control, PhD Thesis, Massey University, New Zealand.
- Russell, N.T. & Bakker, H.H.C. (1997). Modular modelling of an evaporator for long-range prediction. *Artificial Intelligence in Engineering*, **11**, 347-355.

Modular neural network modelling for long-range prediction of an evaporator

- Su, H.-T., McAvoy, J. & Werbos, P. (1992). Long-term predictions of chemical processes using recurrent neural networks: A parallel training approach. *Industrial Engineering Chemistry Research*, **31**, 1338-1352.
- Werbos, P.J. (1990). Backpropagation through time: what it is and how to do it. *Proc. of the IEEE*, **78**, 1550-1560.
- Williams, R.J. & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, **1**, 270-280.
- Willis, M.J., Di Massimo, C., Montague, G.A., Tham, M.T. & Morris, A.J. (1991). Artificial neural networks in process engineering. *IEE Proceedings - D, Control Theory Appl.*, **138**, 256-266.

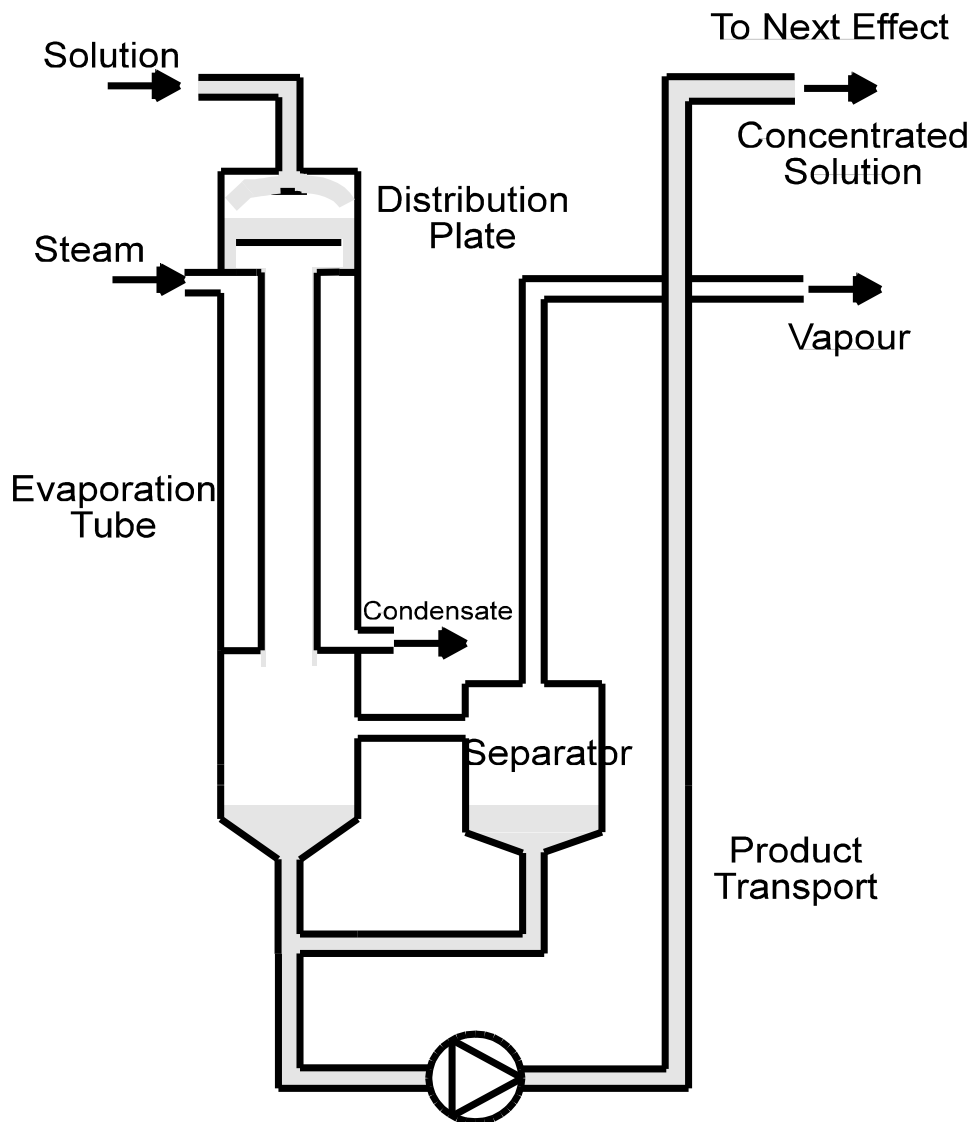


Fig. 1. A single effect of the three-effect evaporator

Modular neural network modelling for long-range prediction of an evaporator

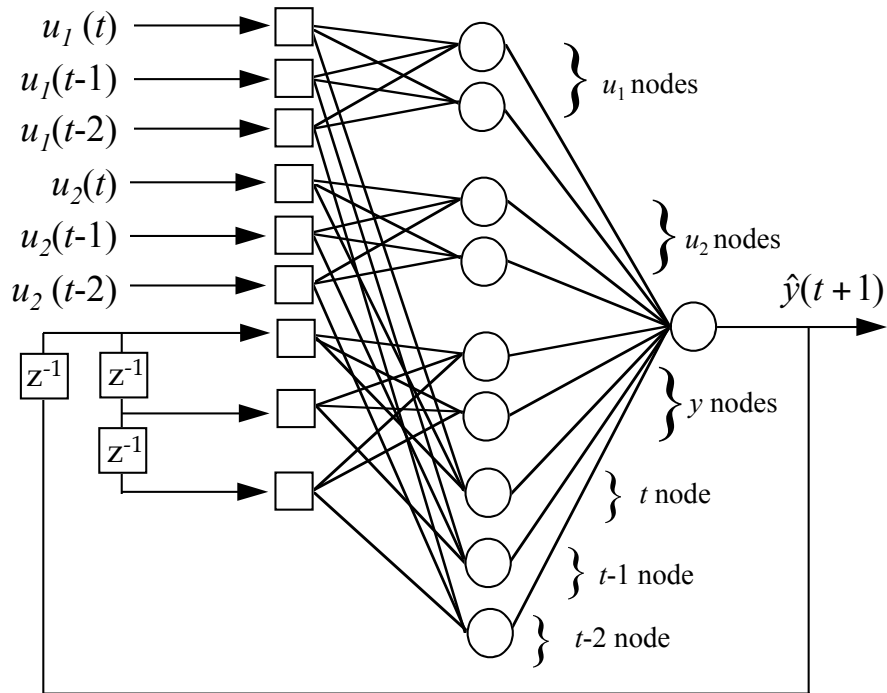


Fig. 2. An externally-recurrent sub-network with localised computation

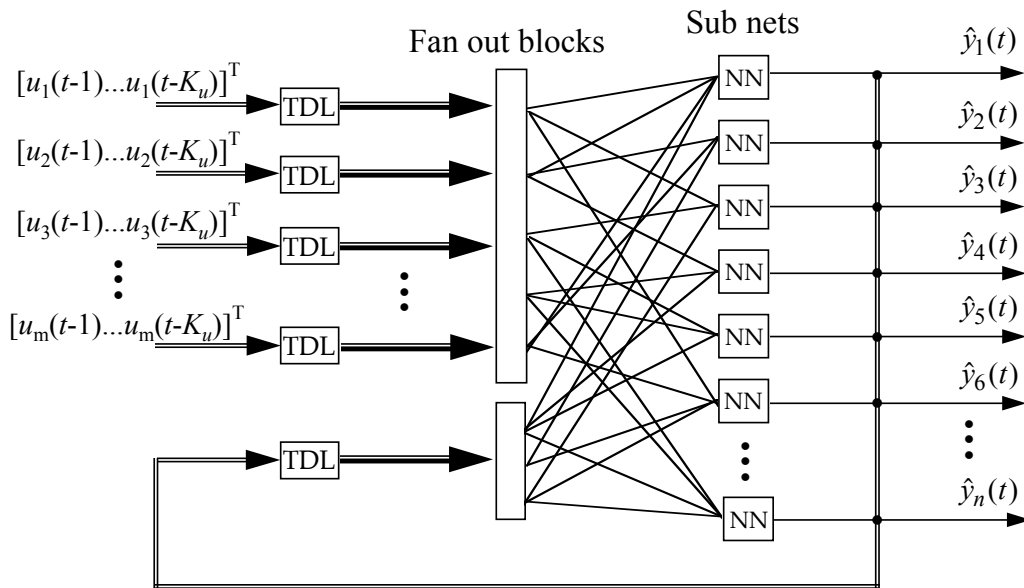


Fig. 3. General structure of the NN evaporator model

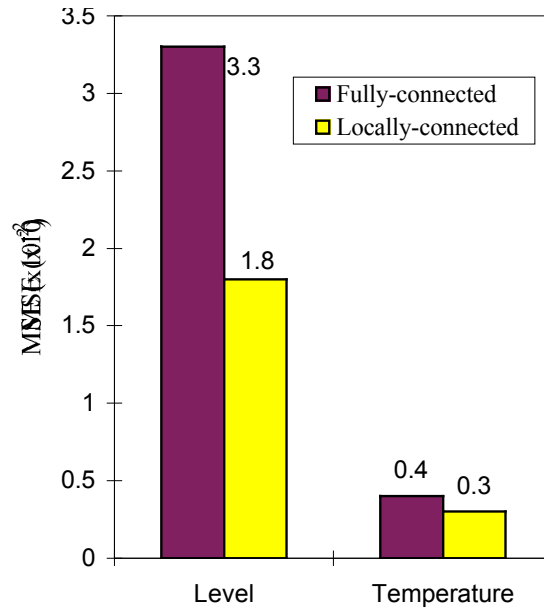


Fig. 4. Long-range prediction errors for the second effect sub-networks

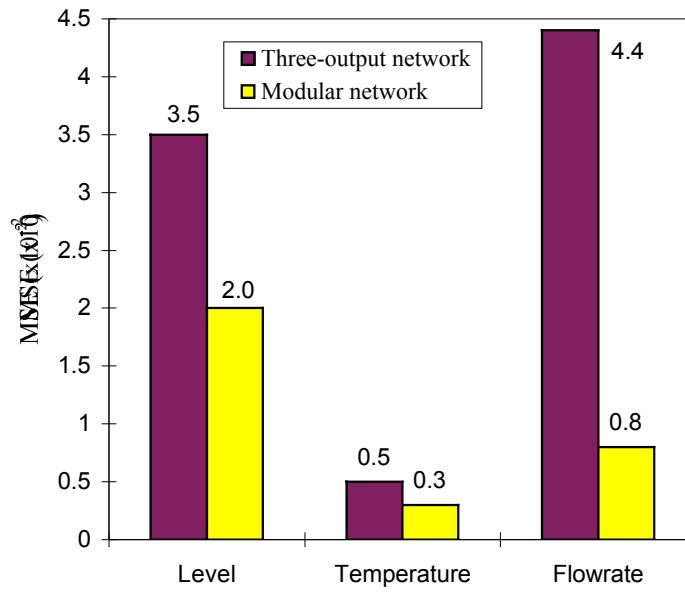


Fig. 5. Long-range prediction errors for the second effect NN models

Modular neural network modelling for long-range prediction of an evaporator

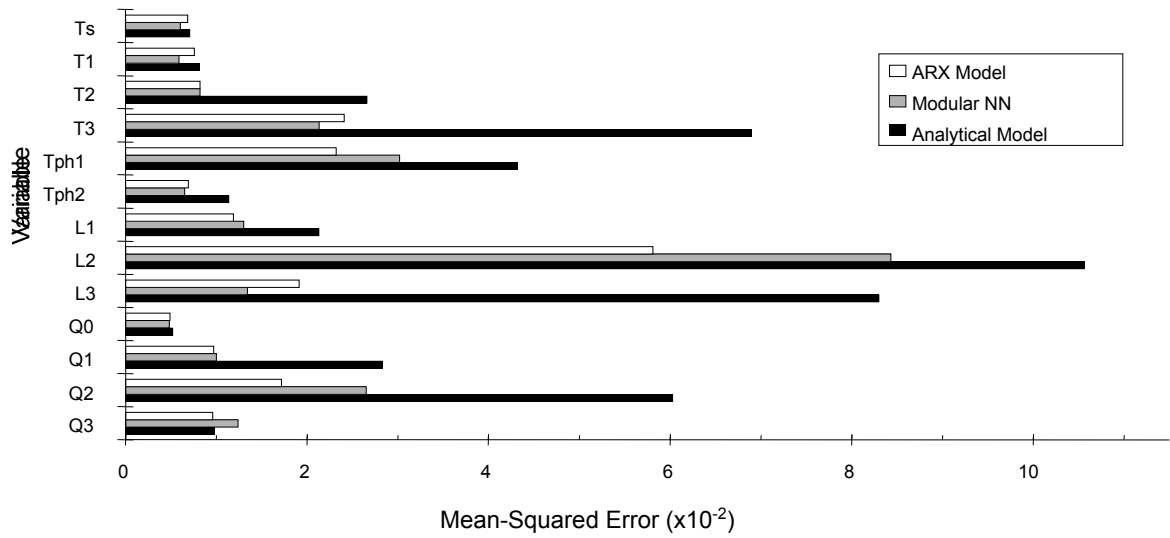


Fig. 6. Comparison of prediction errors for the ARX, NN and analytical models

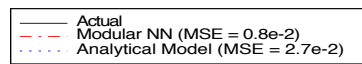


Fig. 7. Comparison of actual data with NN and analytical model predictions for the second effect temperature (T_2)

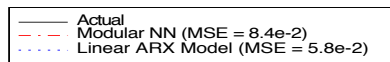


Fig. 8. Comparison of actual data with NN and ARX model predictions for the second effect concentrate level (L_2)

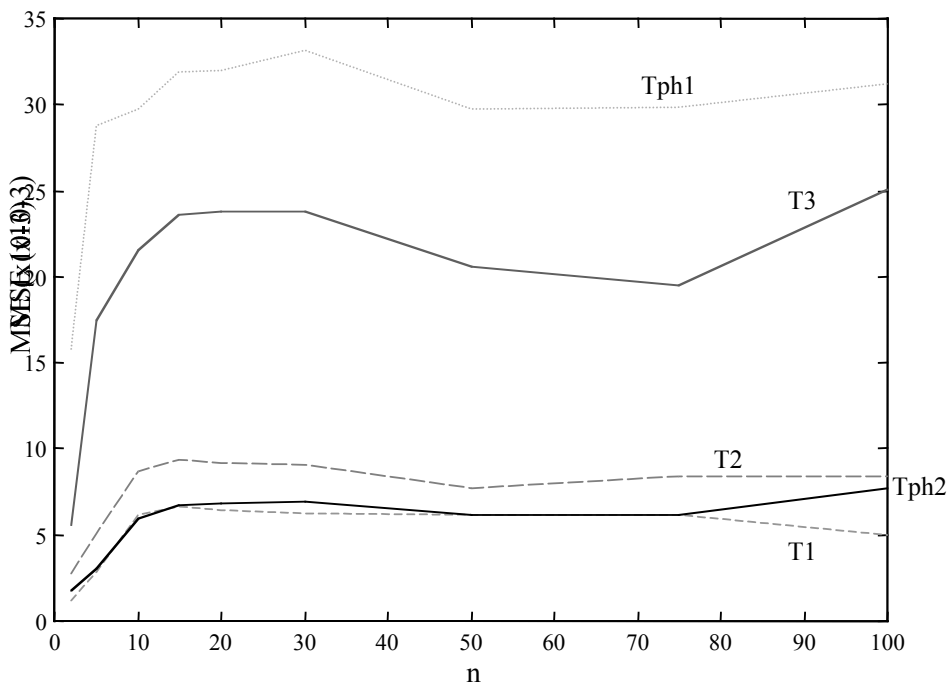


Fig. 9. MSE vs prediction horizon for the evaporator temperatures

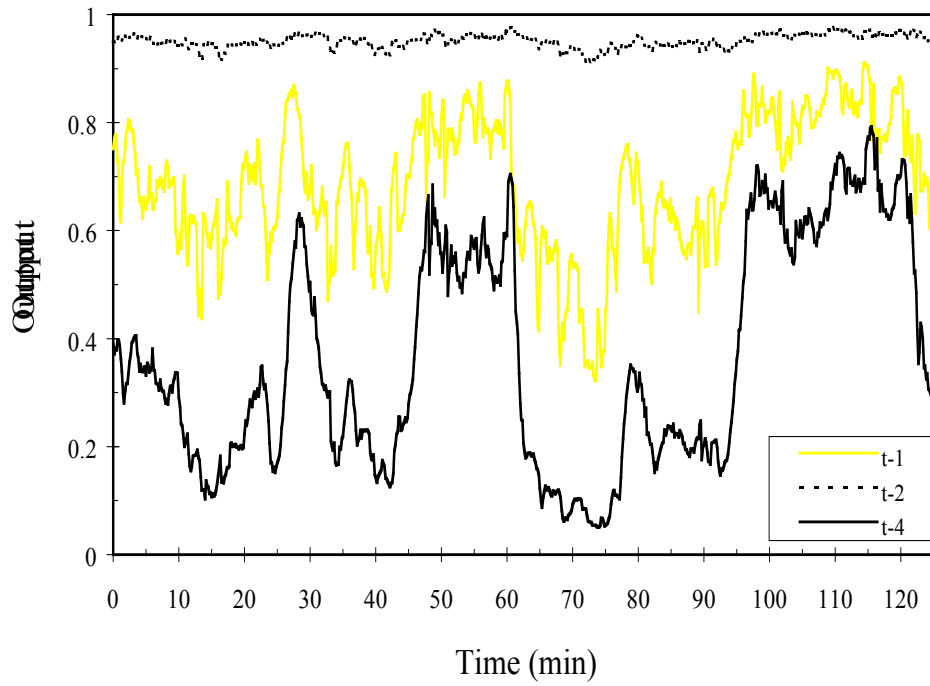


Fig. 10. Outputs from the three time-input hidden neurons

Table 1. Number of parameter for second effect sub-networks

Output	Locally-connected	Fully-connected
T_2	59	57
L_2	59	57

Table 2: Number of parameters in the Effect 2 models

Modular network	Locally-connected 3-output network
137	88

Modular neural network modelling for long-range prediction of an evaporator

Russell, NT

2000-01
