International Journal of Production Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tprs20

Static neural network process models: Considerations and case studies

D. W. Coit , B.T. Jackson & A.E. Smith
Published online: 15 Nov 2010.

To cite this article: D. W. Coit , B.T. Jackson & A.E. Smith (1998) Static neural network process models: Considerations and case studies, International Journal of Production Research, 36:11, 2953-2967, DOI: 10.1080/002075498192229

To link to this article: http://dx.doi.org/10.1080/002075498192229

INT. J. PROD. RES., 1998, VOL. 36, NO. 11, 2953–2967

Static neural network process models: considerations and case studies

D. W. COIT†, B. T. JACKSON‡ and A. E. SMITH‡*

Neural networks are beginning to be used for the modelling of complex manufacturing processes, usually for process and quality control. Often these models are used to identify optimal process settings. Since a neural network is an empirical model, it is highly dependent on the data used in construction and validation. Using data directly from production ensures availability and fidelity, however, the samples may not reflect the entire range of probable operation and, in particular, may not include the optimal process settings. Supplementing production data with observations gathered from designed experiments alleviates the problem of overly focused or incomplete production data sets. This paper considers practical aspects of building and validating neural network models of manufacturing processes, and illustrates the recommended approaches with two diverse case studies.

1. Introduction

Neural networks have many attractive properties for the modelling of complex production systems: universal function approximation capability (Gong 1986, Fahlman and Lebiere 1990), resistance to noisy or missing data, accommodation of multiple non-linear variables with unknown interactions, and good generalization ability. For manufacturing processes where no satisfactory analytic model exists or where a low-order empirical polynomial models is inappropriate, neural networks are a good alternative approach. Recent overviews of neural network applications in manufacturing were compiled by Udo (1992) and Zhang and Huang (1995) who cite such diverse venues as milling, metal cutting, injection moulding, arc welding and spray painting, among others. Other manufacturing process modelling applications can be found in (Andersen *et al.* 1990, Sathyanarayanan *et al.* 1992, Hou and Lin 1993, Smith 1993, Wang *et al.* 1993, Lampinen and Taipale 1994, Yang *et al.* 1994). One motivation for development of neural network process models is that they do not depend on simplified assumptions such as linear behaviour or production heuristics. If a trained neural network model represents the process with fidelity, the neural network can be used to optimize controllable process parameters with regard to some cost metric that measures final product quality, rework and scrapping, labour during production, or use of raw materials. Optimization takes place by searching over controllable (control) variables for the combination of settings that yields the best performance on the cost function. This can be done by fixing the uncontrollable (state) variables at their value or range for the particular time, day, product design, machine and/or operator that are the expected conditions. Alternatively, optimization could be done by letting the state variables become a

Received November 1997.

† Department of Industrial Engineering, 96 Frelinghuysen Road, Rutgers University, Piscataway, NJ 08854-8018 USA.

‡ Department of Industrial Engineering, 1031 Benedum Hall, University of Pittsburgh, Pittsburgh, PA 15261, USA.

* To whom correspondence should be addressed.

search dimension so that their values complement the optimized values of the control variables. This approach, while it might yield a better process outcome, assumes some degree of control over the state variables. Note that the process models referenced and discussed in this paper are static models, that is, there is no explicit effort to capture the dynamics of the process. Static models are applicable to processes where the settings for each piece (or lot) are determined up front, and are not altered for that piece (or lot) using feedback during the process.

A neural network predictive model has the advantages listed above; however, there are drawbacks as well. A primary concern is that neural networks usually act as 'black boxes', and moreover, are empirical black boxes. While process knowledge and physical relationships may guide the formulation of the neural network model, most final models are too complex to explicitly interpret the model components, e.g. interpreting network weighted connections as is done with interpreting slope and intercept coefficients in regression models. To assure confidence in a black box model, the model must be constructed and validated using data which adequately and accurately reflects the process domain. This domain will often include settings and process interactions not commonly seen during production (novel, rare or extreme values), but which it is imperative that the neural model properly handle. This is especially important when the model is used for process optimization, as it is possible that the best combination of process and/or state variable settings will not be found among existing process data. This is also important for process assurance, where the network needs to react properly to uncommon situations to move the process towards normal operating conditions. In such cases, it is proposed that the data used to construct the model should include the results obtained from design of experiments (DOE) along with production data.

This paper is aimed at the technology transfer aspects of neural networks to manufacturing process modelling and optimization by focusing on two highly non-linear processes where there are many variables which affect the ultimate outcome. These are wave soldering of printed circuit boards (PCBs) and slip casting of large ceramic products. Both processes had been controlled using simple regression models and heuristics, yet the manufacturers experienced undesirable levels of rework. Optimizing the processes by trial and error was impractical due to resource and time constraints, so neural models were developed to be used in process optimization. After a discussion of important considerations for building useful neural network process models, the balance of the paper focuses on these two case studies with attention to the interleaving of production data with experimental data.

2. Considerations of neural manufacturing process models

The first step in building manufacturing models is to identify the outcome variable(s) of interest. These may be final or intermediate measures of the product, such as tolerance adherence, or attribute or variable defect measurements. The outcome variables may also be surrogates for the cost metrics of interest, that is, good indicators of the cost metric and more expedient to collect. The surrogate approach was used in both case studies of this paper for reasons discussed later. Once outcome variables have been identified, the process and state variables affecting them must be identified. These may be numerous and will usually fall into general categories of process or machinery settings, ambient conditions, workforce variables, raw material properties, temporal or sequential considerations, and product design aspects. The modeller will want to include those variables which most impact the process as it may not be feasible to include all variables, and if data integrity is in doubt, additional variables may cause erroneous neural predictions or classifications.

Obviously of concern is the size of the data set. More data are beneficial to neural models, as the data are used to both construct and validate the model prior to operation in the production environment. How much data is adequate to model a process is an extremely difficult and problem-specific dilemma. The information imparted to the empirical model by each additional observation (each increase in sample size) is generally monotonically diminishing if the data are sampled uniformly from a population. To minimize the uncertainty in the prediction model, the information content of the data set used to construct the model should be maximized. This implies that data which are non-redundant and which are relevant to the process prediction will impart the most additional information. A designed experiment can be used to choose data which are most beneficial from an information theoretic perspective. There have been a number of discussions on the subject of information theory and neural network training in the literature; Klir and Folger (1988) provide an excellent introduction to information theory and specific aspects are contained in Ahmad and Tesauro (1989), Mhaskar (1993) and Ratsaby and Meir (1996).

Often, it is not possible to collect large data sets because of the cost and time required. If the data set is sparse relative to the number of input and output variables and the complexity of the process to be modelled, special attention must be paid to model training and validation. Overfit is often a symptom of a small data set (White 1989, Geman *et al.* 1992) and can be mitigated by terminating training before the error on the training set is minimized (Twomey and Smith 1993). Validation of models built using small data sets can also be improved by exerting more computational effort and performing a data set resampling, such as cross validation, jackknife or bootstrap (Efron 1982, Gong 1986, Twomey and Smith 1998). With these methods, the accuracy of both the network model and the validation are improved because the entire data set is used for *both* training and testing. These resampling validation methods are beginning to be used in the neural network community (Weiss and Kapouleas 1989, Huston *et al.* 1994, Twomey *et al.* 1995) and may be especially applicable to manufacturing process modelling where data is often costly to obtain. A resampling validation method was used in the first case study in this paper due to the relatively small data set.

2.1. *Aspects of production data*

There are attributes of production data which are attractive. First, it is almost always available in some form, and often in generous quantities. Production data is usually gathered as a matter-of-course, so additional data gathering costs are avoided. Production data includes the effects of the actual production environment. There is no bias towards certain operators, certain procedures or certain machines. For better or worse, the production data includes all the variances encountered during manufacture, ensuring fidelity with operational conditions.

The attributes just cited have their corresponding drawbacks. First, the data may be gathered without rigour, thus including incorrect, erroneous or inexact observations. Some of these may be spotted during an analysis for outliers, but more often it will be impossible to distinguish 'good' observations from 'bad' in the mass of multivariate production data. Second, the production data reflects the process as experienced. That is, the products, settings, workers, machines, ambient conditions, etc.

are those most often encountered during the immediate past. If the production environment has long term shifts, or might experience future changes, production data alone will be inadequate.

2.2. *Aspects of designed data*

Designed data is not simulated or fabricated data; it is actual observation data, but collected under prespecified conditions. When an experiment is designed and performed, the data collected is much more tightly controlled. Exact levels of the input variables (the treatments) are stated, experiments can be blocked to avoid confounding, experiments may be replicated (tests of repeatability) and the sampling distribution is controlled by the designer. Choosing an experimental design to accommodate constraints on sample size and experimental conditions, while also considering interactions, replications and non-linearities, is an important first step. Experimental design for neural network modelling has not been the explicit subject of many papers in the literature; however, a Latin hypercube approach is described by Lunani *et al.* (1995). Experimental data is often collected with more care than production data. Only a single operator, machine or ambient condition may participate. Special tests or measurements might be made which would be impractical during production. Data can be analysed during the course of the experiments to modify the experimental design or the data collection procedures, as needed.

Again, these attributes have undesirable aspects as well. First, designed experiments are usually costly and time consuming. If they are done on production lines, the lines will not be available to produce product. The products manufactured during experimentation will often be scrapped. The conditions may substantially differ from production. For example, tests done in a laboratory setting may be much less variable or contain certain biases not reflective of the manufacturing setting.

2.3. *Combining production data with experimental data in neural networks*

To mitigate the problems describe above, a combined data set is recommended. The production data can be first assessed for representation of past and expected future values. If the production data set is deficit (few or no observations) with regard to certain ranges or interactions, this deficiency can be corrected with experimental data. If the repeatability of the process is unknown, this can be explicitly studied with designed experimentation. It is desirable to characterize not only the variability of the outcome cost measure due to changes in the process, but also the natural variability of the process. Adding experimental data will increase the data set size, but more importantly, will increase the information content of the data set. Extrapolation with any empirical model is imprudent, so use of the model should be confined to data ranges used in model construction and validation.

A possible result of combining production data with experimental data will be an unbalanced training set. For example, if 1000 production observations are available and 50 designed points are taken, the production data may overwhelm the designed data during neural network training. This will especially happen when using a global training procedure such as backpropagation (Werbos 1974, Rumelhart *et al.* 1986) where all trained weights are affected by all training samples and each training observation is (normally) given equal probability mass. A supervised training algorithm that is local, such as radial basis function networks (Moody and Darken, 1989, Poggio and Girosi 1990, Park and Sandberg 1981, Leonard *et al.* 1992) or cascade correlation networks (Fahlman and Labiere 1990) or a self organizing approach such as counter-propagation networks (Whittaker and Cook 1995), will be less susceptible to dominance by the production data. When using global training, the dominance can be eliminated by artificially inflating the probability mass placed on the rarer observations either through:

(1) Direct copying of rare observations to increase their numbers in the training set when using order based training presentation;
(2) Using a probabilistic selection and presentation for training and increasing the probability mass placed on rarer observations (Derouin and Brown 1991) or;
(3) Copying the rare observations and also introducing a small amount of white (Gaussian) noise into the output and/or input(s). This procedure has been sometimes noted to improve the generalization ability and robustness of the final model (Minnix 1991).

There have also been many adjustments to the learning process of supervised neural networks to improve the models in certain circumstances, such as iterative training (Sun *et al.* 1995) or backpropagation speed-ups (Piramuthu *et al.* 1993). These improvements primarily focus on reducing the computational time needed to train a backpropagation network (sometimes at the expense of accuracy in the final network); since training time was not an issue for the case studies of this paper, a conservative and conventional training approach was used. An additional consideration in these case studies was the ease with which the trained network could be translated to compiled source code for use in the plants, further motivating the use of the straightforward backpropagation algorithm.

The next two sections of the paper detail case studies on wave soldering of PCBs and slip casting of ceramics that use the approaches and procedures described. Though each manufacturing process and data characteristics are quite different from the other, the generic approach is appropriate to both. Differences in the experimental facilities, the available production data and the anticipated complexity of the process model were accommodated by customized experimental designs for each manufacturer. The deficits in production data and the values for the designed experiments were identified using the expert knowledge of the process engineers in each case study. These projects have been completed and the resulting neural process control systems have been implemented for use by plant personnel.

3. A wave soldering case study

The wave solder process involves (1) fluxing, (2) preheating, (3) soldering using a wave of solder, (4) cleaning and (5) quality control as shown in figure 1. The process must be adapted according to the PCB design (mass, size, component density, component type, etc.) to optimize quality, viz. minimize solder connection defects. Primary controllable process parameters are the preheater temperatures and the conveyor belt speed. Circuit card manufacturers produce products of great diversity in small lot sizes, compounding the selection of good process settings by trial and error, or by using linear models to select process settings based on predicted average surface temperature of the circuit card (Scheuhing and Cascini 1990, Brinkley 1993). These models can work well, but are limited by their assumption of simple functional form. Malave *et al.* (1992) applied a neural network approach to wave soldering by using circuit card design characteristics as input variables and preheater tempera-

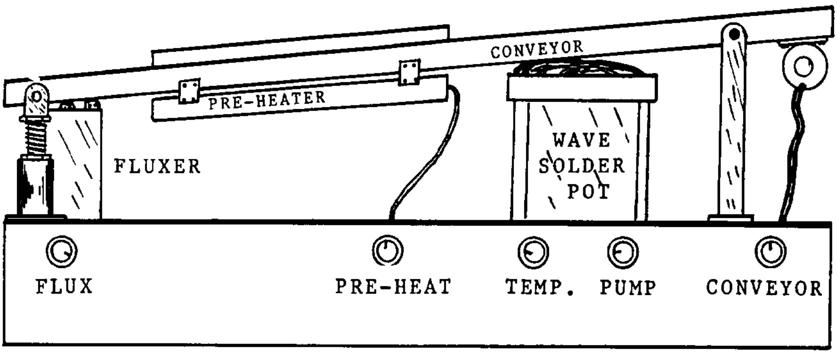

Figure 1. Wave solder process.

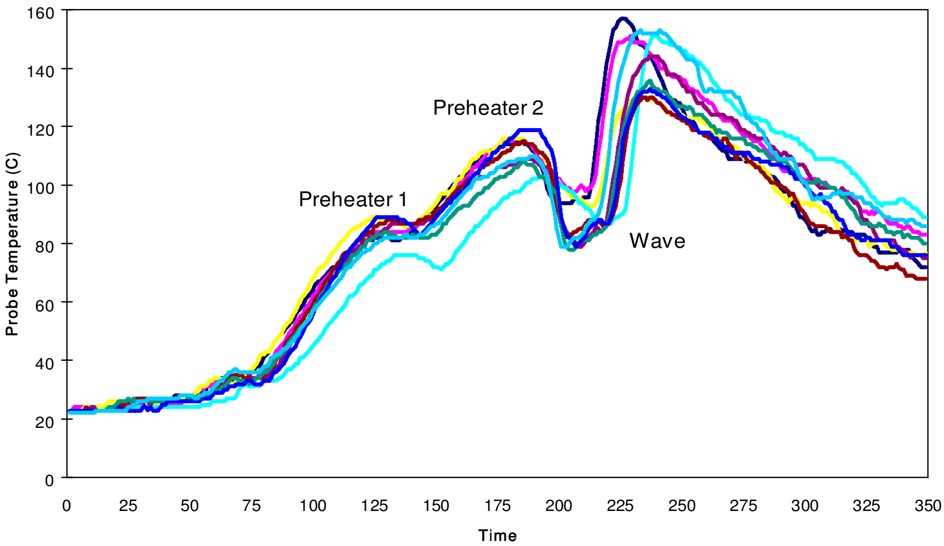

Figure 2. Typical thermal profile of the wave soldering process.

tures and line speed as the output variables. This implicitly assumes that the current process settings are optimal and they were unable to achieve successful results.

The approach taken in this case was to model the single most important determinant of soldering quality – summary of thermal condition (mean and standard deviation of surface temperature, and mean rate of change of surface temperature) of the PCB as it enters the solder wave. Thus, known physical characteristics of the process were incorporated into the model by the selection of the thermal variables to be modelled. Thermal condition at the wave is only observable through special experimentation, as described below, and it is not feasible to measure the thermal characteristics of each PCB during production. To measure thermal condition, temperature probes were attached on top of the PCB, and fed into two MOLE (multichannel occurrent logger evaluator) data recording devices. This setup allowed accurate sampling at 1 Hertz of 8 to 10 temperatures at distinct locations on the PCB. Figure 2 presents a typical thermal profile of the wave soldering process, showing the temperature of each probe as the PCB travels through fluxing, preheating and the solder wave. The temperatures were dependent not only on the wave solder process settings (preheater temperatures and conveyor belt speed), but also on the design characteristics of the PCB itself (e.g., size, thickness, type and distribution of component). Therefore, the choice of probe locations was customized for each PCB design so that both average temperature and temperature extremes, such as those encountered near a heat sink, or near a large component, could be properly characterized.

3.1. *Production data*

In this case study, even the production data had to be gathered especially for the model. This is because detailed thermal condition of the circuit cards requires the special testing apparatus described above, and cannot be done on an ongoing basis. However, these special measurements were incorporated for a two month period into regular production, where one or two PCBs per lot were instrumented and measured.

Using the experimental setup described, the manufacturer gathered 44 production observations of thermal production data showing the design specifications of each PCB, the settings of the four preheaters and the line speed. Because of the small lot sizes, a good variety of PCB designs were measured. Missing from the production data was the effect of alterations in preheater temperatures and line speed for a given PCB design, and an assessment of natural process variability.

3.2. *Experimental data*

For the wave soldering process, the independent variables can be classified as card design parameters or process settings. The process parameters (conveyor belt speed and four preheater temperatures) could be fully and independently adjusted over the range of possible process levels. It was not, however, as easy to adjust the design variables with the same degree of latitude. For example, the number of ground and power planes and the cumulative component mass are both variables which will affect PCB surface temperature, but they could not be independently adjusted because there was a small number of PCB designs available for testing. Therefore, the DOE was concerned with altering the five process settings only for two typical PCB designs. Conserving the number of experimental runs was important because the experiments were conducted on the actual production equipment, which precluded production during the experiments. A benefit of this setup was that the conditions observed and personnel involved during experimentation were the same as those which would be experienced during production.

It was important to investigate non-linear effects, but there was an insufficient number of test cards to support a comprehensive three-way or four-way fractional factorial design. As a compromise, a composite star experimental design with fewer test trials (56) was chosen as presented in table 1. As indicated in the table the experimental runs were randomized. This design collects data at five levels for each independent variable and allows for the characterization of linear effects, linear interactions and quadratic effects, but will not model any interaction effects involving the quadratic terms. Two typical, but dissimilar PCBs were selected to characterize the design limits of the production cards. The complete experimental design was performed separately for each PCB. The only replication occurred at the origin (all factors set to 0 level); however previous experiments had indicated that the process was very repeatable, i.e. had minimal natural variability.

Preheat 1	Preheat 2	Preheat 3	Preheat 4	Speed	Run
- 1.664	0	0	0	0	26
- 1	- 1	- 1	1	- 1	5
- 1	- 1	- 1	- 1	1	19
- 1	- 1	1	1	1	21
- 1	- 1	1	- 1	- 1	23
- 1	1	- 1	- 1	- 1	7
- 1	1	- 1	1	1	16
- 1	1	1	1	- 1	3
- 1	1	1	- 1	1	4
0	- 1.664	0	0	0	25
0	0	- 1.664	0	0	6
0	0	0	- 1.664	0	22
0	0	0	0	- 1.664	13
0	0	0	0	0	1
0	0	0	0	0	15
0	0	0	0	1.664	11
0	0	0	1.664	0	2
0	0	1.664	0	0	17
0	1.664	0	0	0	9
1	- 1	- 1	1	1	10
1	- 1	- 1	- 1	- 1	24
1	- 1	1	1	- 1	12
1	- 1	1	- 1	1	28
1	1	- 1	- 1	1	8
1	1	- 1	1	- 1	20
1	1	1	- 1	- 1	14
1	1	1	1	1	27
1.664	0	0	0	0	18

Preheat:	- 1.664 = 400° F	Speed:	- 1.664 = 1.4 ft/m
	- 1 = 460° F		- 1 = 1.8 ft/m
	0 = 550° F		0 = 2.2 ft/m
	1 = 640° F		1 = 2.6 ft/m
	1.664 = 700° F		1.664 = 3.0 ft/m

Table 1. Experimental design for the wave soldering process.

3.3. *Wave solder neural models*

A combined data set of production data ($n = 44$) and the DOE data ($n = 56$) was used for neural network development. Since the number of observations of each data set were similar (44 and 56), no compensation was needed to eliminate the effects of one data set overwhelming the other during backpropagation training. A set of neural networks was trained to predict the temperature condition of the circuit card at the banks of preheaters and at the solder wave as a function of fourteen inputs that described card design parameters (length, width, thickness, card mass, component mass, number of layers, ground planes and connections, presence or absence of a heat sink) and process settings (four preheater temperatures and conveyor belt speed). The temperature condition was characterized by mean temperature, standard deviation and rate of change (temperature gradient) at the wave. Separate models were developed to predict each of these temperature condition metrics (mean, standard deviation, gradient) calculated over the 8 to 10 probes (described in more detail in Coit *et al.* 1994).

Since the combined data set was still relatively small, a resampling approach to training and validation was adopted. A five fold cross validation (also called the jackknife validation method) of the 100 observations of the combined data was used. Grouped cross validation divides the available data n into k groups, each of size n/k. k models are then constructed, each using all but one of the data groups for model construction, and the held out group for the kth model is used for validation. A final model, which is used for application, is built using all the data. Prediction error of the final model is estimated using the mean of testing set errors of the k grouped cross validation models. Grouped cross validation uses all available data for both model construction and model validation, but requires the construction of $k + 1$ models, i.e. training $k + 1$ neural networks. The error is estimated as the mean squared error over all k networks for each point in the testing set as shown in equation 1:

$$\widehat{Error}_{\mathrm{GCV}} = \frac{1}{n}\sum_{l=1}^{k}\sum_{j=1}^{n/k}\left[y_{lj} - \hat{f}\left[T_{(k)}, x_{lj}\right]\right]^2, \tag{1}$$

where $\hat{f}\left[T, x\right]$ is the neural network output for input vector $\mathbf{x}$ and y is the actual output. Using this approach, five neural networks were trained, each using 4/5 of the data set for training and 1/5 for testing. A different 1/5 for testing was used for each cross validation network. Then a final neural model was built using all 100 observations for training. This final model is actually used in production, while the five cross validation networks are used to estimate the prediction error of this final network (which cannot be calculated directly because all observations are used in its construction). Therefore, a total of 18 networks were developed; five cross validation networks plus one final network for three different outputs (mean, standard deviation and rate of change of surface temperature). An ordinary backpropagation algorithm with a smoothing factor (equation 2) and a unipolar sigmoid transfer function was used for all networks, which were fully connected with one hidden layer.

$$\Delta w_{t,i} = \eta(\alpha \Delta w_{t-1,i} + (1 - \alpha) o_i \delta_i), \tag{2}$$

where Δw_i is the change in weight connecting to neuron i, o_i is the output of neuron i, δ_i is the error of the output of neuron i times the derivative of the sigmoid function, η is the training rate (0.10) and α is the smoothing term (0.90) for training iteration t.

Training terminated when the marginal improvement in the error over the training set became close to zero. The number of epochs required to reach this convergence point varied with the network, ranging from 200 to 7700. A variety of network architectures and training parameters were tested, before the settings were fixed. Network training was not sensitive to learning rate or exact termination point, however, it was somewhat sensitive to the number of hidden neurons.

The network architectures and results are shown in table 2. It can be seen that the networks do a good job of predicting thermal condition, especially mean card temperature. When all three predictions were combined, a good categorical assessment of solder quality could be made. Another neural network was trained to take the prediction of the three thermal indicators at the wave and predict category of solder quality (excellent, good, fair). This final neural network performed at a rate of 82.6% correct predictions for the test set.

Output	Training set MAE	Training set % MAE	Test set MAE	Test set % MAE	Network architecture
Mean temperature	5.22	5.57	6.63	7.19	14, 10, 1
Temperature gradient*	0.2397	17.55	0.3176	24.02	14, 9, 1
Temperature standard deviation	1.92	20.95	2.44	27.26	14, 6, 1

* In °C second.

Table 2. Performance of predicting PCB thermal condition averaged over five cross validation networks.

4. A slip casting case study

In contrast to the wave solder process, slip casting of large ceramic pieces is labour intensive with many controllable and uncontrollable variables. Product design does not change frequently, but process conditions can change substantially from day to day. The manufacturing steps consist of the following: (1) preparation of slip, (2) casting slip in a mould, (3) drying the slip and removing the mould, (4) air drying of cast piece (5) spray glazing the dried product, (6) firing the glazed product and (7) inspection of the finished product. Step 2 of the process is slip casting, where a suspension (the slip) is poured into a mould and the liquid phase is separated by capillary phenomena, leaving a solid piece that takes the shape of the mould (Lambe 1958). The primary causal factor for cast fractures and/or deformities is the distribution of moisture content inside the cast before firing in a kiln. When the moisture differential, or moisture gradient, inside the wall of the cast is too steep, it results in stress differences that cause the piece to deform and eventually fracture. In order to have a good cast, and therefore a solid, durable product, the moisture gradient should be as uniform as possible. Another important output measure is cast rate, which is actually the thickness of the cast achieved during the time the slip is in the mould. A larger cast rate will result in more efficient production, as cast time decreases.

The quality of the cast and the cast rate depend on the chemical properties of the slip, the ambient conditions in the plant and the mould conditions. Ceramic engineers run a series of tests that emulate the behaviour of the slip during casting. On the basis of these tests the engineers can modify the slip's composition to produce a 'forgiving' slip. An ideal forgiving slip compensates for the effects of other, less controllable, variables involved in slip casting including ambient conditions and mould conditions. The manufacturer in this case study routinely measured ten slip property variables, two ambient variables, time in the cast and two outcome variables. These are named and defined in table 3.

Although there have been a few computer-aided improvements in the slip casting process, such as an expert system aimed at slip particle effects (Dinger 1990) it still remains basically an art. This is because there are no satisfactory analytic descriptions of casting dynamics. It is affected by many human and non-human variables, and the effect of the interdependencies of these variables are only manifested at the end of the process, after the firing of the cast. This provided the motivation for the project to develop a predictive model of the slip casting process which would allow the manufacturer to optimize controllable process parameters without wasteful and time consuming test casts.

Parameter	What does it measure?
Cast time	The time the liquid slip is left in the mould before draining.
Slip temperature	The temperature of the slip.
Brookfield – 10 rpm	Viscosity of the slip at 10 revolutions per minute.
Brookfield – 100 rpm	Viscosity of the slip at 100 revolutions per minute.
Initial reading	Initial slip viscosity (taken at $3\frac{1}{2}$ minutes).
Build up	Change in slip viscosity from initial reading (taken after 18 minutes).
Filtrate rate	The rate at which the slip filtrates.
Gelation	Thixotropy (viscosity versus time).
Slip cake weight	Approximation of the cast rate without considering a mould.
Cake weight water retention	Moisture content of the cake. See slip cake weight.
Sulfate	Proportion of soluble sulfates in the slip.
Temperature	The temperature of the plant.
Relative humidity %	The humidity level of the plant.
Cast rate	Inches of cast build up during a specified time interval, 50–80 min.
Moisture gradient	Quantitative measure of the moisture differential in a cast wall.

Table 3. Slip casting process parameters.

4.1. *Production data*

The manufacturer had over two years of daily production describing the parameters of table 3 for the quality metric of cast rate, and nearly one year of the same data with the quality metric of moisture gradient. Therefore, the production data set was relatively large and formed the sole basis of preliminary neural models as described in Martinez *et al.* (1994). Although a variety of process settings was included in these data sets, extreme values of three key parameters were missing. Two were extreme ambient conditions (very high and very low temperatures, and very high and very low humidities), both state variables. The third was the sulfate content in the slip, which is adjusted daily to achieve the desired 'forgiving' slip. In order to optimize this process, which is substantially dependent on these three variables, a neural model which could operate effectively over an expanded range of values was required.

4.2. *Experimental data*

To explore the extreme ranges of sulfate, plant temperature and plant humidity, designed experiments were performed. This experimentation was also used to gauge the effect of mould condition on the cast. Whether the mould is relatively wet (old) or relatively dry (new) will affect the cast. Since the manufacturer did not gather production data on mould condition, this factor could only be explored through experimentation. For the slip casting operation, a traditional fractional factorial design, presented in table 4, was developed with five factors at two levels. The main concern was to conserve test casts, as they were quite expensive to develop and had to be scrapped. This design is a $2^{\mathrm{V-I}}$ design with the ability to model two-way and three-way interactions, but no non-linear terms. The experimental constraints differed from the wave soldering case study, as here the experimental products were produced by personnel not normally involved in daily production, and involved pouring the slip from buckets rather than the hoses which are used in production. Therefore, there could be confounding of the results since these conditions were somewhat dissimilar from production conditions.

Day number	Sulfate content	Relative humidity	Plant temp.	Cast time	Mould condition	Run number
1	+	+	+	+	+	5
1	+	+	+	-	+	3
1	+	+	-	+	-	4
1	+	+	-	-	-	2
1	+	-	+	+	-	8
1	+	-	+	-	-	7
1	+	-	-	+	+	1
1	+	-	-	-	+	6
2	-	+	+	+	-	14
2	-	+	+	-	-	16
2	-	+	-	+	+	12
2	-	+	-	-	+	9
2	-	-	+	+	+	10
2	-	-	+	-	+	15
2	-	-	-	+	-	13
2	-	-	-	-	-	11

Table 4. Experimental design for the slip casting process.

The high (+) and low (-) levels were selected to characterize the expanded range over operating conditions found in the plant. As indicated in table 4, the sulfate content is confounded with the day number. This was undesirable but necessary, because it was not possible to concurrently maintain two types of slip given the constraints of the plant. Each run consisted of casting a large piece for a predetermined time, draining the excess slip and then measuring the cast rate and the moisture gradient at three locations of each piece and averaging them. Two replications were performed for each run.

4.3. *Slip casting neural models*

Two neural networks were developed to predict the cast rate and the moisture gradient as a function of the variables listed in table 3. Since the data sets were large ($n = 1000$ for cast rate and $n = 350$ for moisture gradient), a resampling approach was not needed, and the more common random splitting of the data into a testing set (20%) and a training set (80%) was used. The experimental data set was small relative to the production data set, so it might seem that some artificial increase in the sampling weight of the experimental data was needed. However, the experimental data was taken only from the *very* extreme ranges of the variables studied. Therefore, the only information the neural models had on the output variables' behaviour in these extreme ranges was from the experimental data, thus artificial inflation of its numbers during training was not needed.

All networks were developed and trained in the same manner as in the wave soldering project, the only difference being the network architectures and the learning rate, η which was increased to 0.30. The cast rate neural network had thirteen inputs (ten slip variables, two ambient variables and cast time), two hidden layers, with eight hidden neurons in each layer, and a single output. Training converged after 2600 epochs. It was able to predict the cast rate with a normalized mean absolute error of 3.20% on the test set of 200 observations. The mean moisture gradient neural model was developed in a similar way using the same input variables, but with a single hidden layer of eight neurons and a smaller combined data set of 350 observations. It converged after 2000 epochs with a normalized mean absolute error of 6.04% on the test set of 70 observations. These two neural networks were used as the foundation for a larger system which selected optimum values of controllable variables given fixed values of the state variables. Optimum is the sense of maximum cast rate, as predicted by one neural network, and minimum moisture gradient, as predicted by the other neural network.

5. Conclusions

When using a neural network to control and optimize a manufacturing process, the integrity and balance of the training and validation data sets dictate the quality of the resultant model. Optimization thoroughly searches the feasible region of allowable process settings, and therefore, unbiased and accurate predictions are essential for combinations of process settings which may not have been regularly encountered during historic operating scenarios. The use of experimental design strategies is often necessary to provide the required complementary data to the production data.

Two diverse manufacturing case studies were presented. While the manufacturing processes and the specific experimental designs were different, the overall modelling procedure was similar. Development of sound models was accomplished by first defining interim surrogates for product quality, and then by determining problem-specific experimental designs. The experimental data was combined with the production data, and neural networks were trained and validated on the combined data set. After careful validation of the prediction accuracy over the entire range of anticipated operating conditions, the final neural network models have been implemented at the manufacturing plants. Assurance and user confidence were vital considerations of implementation because of concerns about using opaque models to support critical decision making. Inappropriate, sub-optimal and incorrect decisions based on flawed neural predictions would not be recognized until the decisions had been implemented and caused undesirable outcomes in the plant or products. This was circumvented by including data from all anticipated ranges of production operation.

Acknowledgments

Part of this work was supported by the Ben Franklin Technology Center of Western Pennsylvania, Lockheed Martin Corporation and the US National Science Foundation grant DMI 9502134 CAREER. The second author acknowledges support of a Graduate Fellowship from the US National Science Foundation.

References

Ahmad, S. and Tesauro, G., 1989, Scaling and generalization in neural networks: a case study. In *Advances in Neural Information Processing Systems*, I (D. S. Touretzky, ed.) (San Mateo, CA: Morgan Kaufmann), pp. 160–168.

Andersen, K., Cook, G. E., Karsai, G. and Ramaswamy, K., 1990, Artificial neural networks applied to arc welding process modeling and control. *IEEE Transactions on Industry Applications*, **26**, 824–830.

Brinkley, P. A., 1993, Northern Telecom achieves improved quality by combining DOE and SPC. *Industrial Engineering*, **25**, 63–65.

Coit, D. W., Billa, J., Leonard, D., Smith, A. E., Clark, W. and El-Jaroudi, A., 1994, Wave soldering process control modeling using a neural network approach. In *Intelligent Engineering Systems Through Artificial Neural Networks*, Vol. 4, C. H.

Dagli, B. R. Fernandez, J. Ghosh and R. T. S. Kumara (eds) (New York: ASME Press), pp. 999–1004.

Derouin, E. and Brown, J., 1991, Neural network training on unequally represented classes. In *Intelligent Engineering Systems Through Artificial Neural Networks*, C. H. Dagli, S. R. T. Kumara and Y. C. Shin (eds) (New York: ASME Press), pp. 135–140.

Dinger, D. R., 1990, Expert systems for use in ceramic processing. *Ceramic Engineering Science Proceedings*, **11**, 320–331.

Efron, B., 1982, *The Jackknife, the Bootstrap and Other Re-sampling Plans* (Philadelphia: SIAM).

Fahlman, S. E. and Lebiere, C., 1990, The cascade-correlation learning architecture. In *Advances in Neural Information Processing Systems*, Vol. 2, D. S. Touretzky (ed.) (San Mateo, CA: Morgan Kaufmann), pp. 524–532.

Funahashi, K.I, 1989, On the approximate realization of continuous mappings by neural networks. *Neural Networks*, **2**, 183–192.

Geman, S., Bienenstock, E. and Doursat, R., 1992, Neural networks and the bias/variance dilemma. *Neural Computation*, **4**, 1–58.

Gong, G., 1986, Cross-validation, the jackknife, and the bootstrap: excess error estimation in forward logistic regression. *Journal of the American Statistical Association*, **81**, 108–113.

Hornik, K., Stinchcombe, M. and White, H., 1989, Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359–366.

Hou, T.-H. and Lin, L., 1993, Manufacturing process monitoring using neural networks. *Computers and Electrical Engineering*, **19**, 129–141.

Huston, T. L., Smith, A. E. and Twomey, J. M., 1994, Artificial neural networks as an aid to medical decision making: comparing a statistical resampling technique with the train-and-test technique for validation of sparse data sets. In *Artificial Intelligence in Medicine: Interpreting Clinical Data*. AAAI Press Technical Report SS-94-01, pp. 70–73.

Klir, G. J. and Folger, T. A., 1988, *Fuzzy Sets, Uncertainty and Information* (Englewood Cliffs, NJ: Prentice Hall).

Lambe, C. M., 1958, Preparation and use of plaster molds. In *Ceramic Fabrication Processes* W. D. Kinger (ed.) (New York: Wiley), chapter 3.

Lampinen, J. and Taipale, O., 1994, Optimization and simulation of quality properties in paper machine with neural networks. *Proceedings of the IEEE International Conference on Neural Networks*, pp. 3812–3815.

Leonard, J. A., Kramer, M. A. and Ungar, L. H., 1992, A neural network architecture that computes its own reliability. *Computers and Chemical Engineering*, **16**, 819–835.

Lunani, M., Sudjianto, A. and Johnston, P. L., 1995, Generating efficient training samples for neural networks using Latin hypercube sampling. In *Intelligent Engineering Systems Through Artificial Neural Networks*, **5**. C. H. Dagli, M. Akay, C. L. P. Chen, B. R. Fernandez and J. Ghosh (eds), (New York: ASME Press), pp. 209–214.

Malave, C. O., Sastri, T. and Johnson, R., 1992, Prediction of wave solder machine parameters based on printed circuit board design characteristics. In *Intelligent Engineering Systems Through Artificial Neural Networks*, Vol. 2, C. H. Dagli, L. I. Burke and Y. C. Shin (eds) (New York: ASME Press), pp. 833–838.

Martinez, S., Smith, A. E. and Bidanda, B., 1994, Reducing waste in casting with a predictive neural model. *Journal of Intelligent Manufacturing*, **5**, 277–286.

Mhaskar, H. N., 1993, Approximation of real functions using neural networks. *Proceedings of the International Conference on Computational Mathematics* (New Delhi: World Scientific Press), pp. 267–268.

Minnix, J. I., 1991, An analysis of the effects of noisy training sets on the fault tolerance of neural networks. *Proceedings of the 1991 IEEE Systems, Man, and Cybernetics Conference*, pp. 713–718.

Moody, J. and Darken, C., 1989, Fast learning in networks of locally tuned processing units. *Neural Computation*, **1**, 281–294.

Park, J. and Sandberg, I. W., 1991, Universal approximation using radial-basis function networks. *Neural Computation*, **3**, 246–257.

Piramuthu, S., Kuan, C.-M. and Shaw, M. J., 1993, Learning algorithms for neural-net decision support. *ORSA Journal on Computing*, **5**, 361–373.

Poggio, T. and Girosi, F., 1990, Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, **247**, 978–982.

Ratsaby, J. and Meir, R., 1996, Finite sample size results for robust model selection; application to neural networks. Technical Report of the Faculty of Electrical Engineering, Haifa University.

Rumelhart, D. E., McClelland, J. L. and the PDP Research Group, 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, **1** (Cambridge, MA: MIT Press).

Sathyanarayanan, G., Lin, I. J. and Chen, M.-K., 1992, Neural network modelling and multiobjective optimization of creep feeding grinding of superalloys. *International Journal of Production Research*, **30**, 2421–2438.

Scheuhing, R. B. and Cascini, R., 1990, Wave solder thermal profiling using multivariate linear regression techniques. *Electronic Manufacturing*, **36**, 20–28.

Smith, A. E., 1993, Predicting product quality with backpropagation: a thermoplastic injection moulding case study. *International Journal of Advanced Manufacturing Technology*, **8**, 252–257.

Sun, X., Golden, B. and Wasil, E., 1995, The fine-tuned learning enhancement to the standard backpropagation algorithm. In *Intelligent Engineering Systems Through Artificial Neural Networks*, Vol. 5, C. H. Dagli, M. Akay, C. L. P. Chen, B. R. Fernandez and J. Ghosh (eds) (New York: ASME Press), pp. 125–131.

Twomey, J. M. and Smith, A. E., 1993, Nonparametric error estimation methods for validating artificial neural networks. In *Intelligent Engineering Systems Through Artificial neural Networks*, Vol. 3, C. H. Dagli, L. I. Burke, B. R. Fernandez, J. Ghosh (eds) (New York: ASME Press), pp. 233–238.

Twomey, J. M. and Smith, A. E., 1998, Bias and variance of validation methods for function approximation neural networks under conditions of sparse data. *IEEE Transactions on Systems, Man, and Cybernetics,* Part C, **28** (3).

Twomey, J. M., Smith, A. E. and Redfern, M. S., 1995, A predictive model for slip resistance using artificial neural networks. *IIE Transactions*, **27**, 374–381.

Udo, G. J., 1992, Neural network applications in manufacturing processes. *Computers and Industrial Engineering*, **23**, 97–100.

Wang, Q., Sun, X., Golden, B. L., DeSilets, L., Wasil, E. A., Luco, S. and Peck, A., 1993, A neural network model for the wire bonding process. *Computers and Operations Research*, **20**, 879–888.

Weiss, S. M. and Kapouleas, I., 1989, An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. *Proceedings of the 11th International Joint Conference on Artificial intelligence*, pp. 781–787.

Werbos, P. J., 1974, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* Thesis, Harvard University.

White, H., 1989, Learning in artificial neural networks: a statistical perspective. *Neural Computation*, **1**, 425–464.

Whittaker, A. D. and Cook, D. F., 1995, Counterpropagation neural network for modelling a continuous correlated process. *International Journal of Production Research*, **33**, 1901–1910.

Yang, Y., Cheng, Y., Zhao, R. and Govind, R., 1994, Process optimization using neural networks. *Proceedings of the IEEE International Conference on Neural Networks*, pp. 4635–4639.

Zhang, H.-C. and Huang, S. H., 1995, Applications of neural networks in manufacturing: a state-of-the-art survey. *International Journal of Production Research*, **33**, 705–728.