Abstract—This paper presents a systematic approach to the complex problem of RBF-ARX modeling. First, we point out that many of the nonlinear features of a time series may be represented by a relatively simple RBF-ARX model. A method for estimating the number of RBF centers is then proposed based on the behavior of the state variable, and initial values for the centers are found. Linear estimation methods are implemented to select the initial lag orders of candidate models. Model parameters are found by nonlinear estimation and candidate models are compared using AIC, SBC criteria and other diagnostic checks. The modeling approach is shown to work well in practice by estimating optimum RBF-ARX models for real and simulated time series data and comparing the results with those of previous authors. Diagnostic checking also confirms the validity of the method.

Key words—nonlinear systems, parameter estimation, radial basis function networks (RBF), RBF-ARX models, initial values, diagnostic tests.

1. INTRODUCTION
Time series models have found many applications in prediction and control, ranging over a variety of fields. However, traditionally, most time series models used in practice have been linear in form. It is only over the past two decades that a range of nonlinear models have been proposed, for example bilinear models (Granger and Anderson, 1978; Subba-Rao and Gabr, 1984) threshold models (Tong, 1983), state-dependent models (Priestley, 1980, 1988), exponential autoregressive models (Ozaki and Oda, 1978; Haggan and Ozaki, 1981). These models too have been applied in many situations where linear models were incapable of representing the nonlinear behavior of the process under study. Vesin (1993) proposed a radial basis function (RBF) approach to autoregressive modeling, and was followed by Shi et al. (1999) who developed the radial basis function autoregressive (RBF-AR) model, which was further modified by Peng et al. (2001, 2003a) to the RBF-ARX model, extending the idea of RBF-AR to the case where there are several external inputs to the system. The RBF-AR and RBF-ARX models have been used with success, including the modeling and control of a thermal power plant (Peng et al., 2002, 2003a, 2003b). However, RBF-ARX models are highly complex, having large numbers of parameters, and the modeling process reported so far has been somewhat opaque. The process consists of finding initial values for the RBF centers, and scaling parameters, and also choosing appropriate lag orders for the model. The optimum model in terms of maximum likelihood is then found by nonlinear optimization. Different models may be compared by using information theoretic criterion such as AIC (Akaike, 1974) or SBC (Schwarz, 1978). However, without guidelines on how the initial values for centers or lag orders may be selected, the practitioner in the field is likely to find the RBF-ARX modeling process rather daunting.

In section 2, we point out that the RBF-ARX model used by Peng et al. (2002, 2003b, 2003c) is unnecessarily complex in form and that comparable or better results may be obtained by using a simpler form of the RBF-ARX model. Section 3 of this paper presents a systematic methodology for choosing the initial values for model centers, and model orders, based on the behavior of the state and an extension of linear modeling techniques. In section 3, we also discuss the diagnostic checking of the fitted model, and emphasize the importance of checking whether the residuals of the models may in fact be regarded as Gaussian white noise.

In section 4, the method is applied to theoretical and practical situations, including the modeling of a thermal power plant previously reported in Peng et al. (2003b, 2003c), and the results obtained are compared with those found by other researchers. We find that the suggested method obtains superior, or comparable, results in terms of
residual variance, AIC and SBC. In particular, the RBF-ARX model suggested in this paper for fitting a model to a thermal power plant is much less complex than that fitted by Peng et al. (2003b, 2003c), but nevertheless performs equally well in times of reduced residual variance and AIC, SBC. We also undertake thorough diagnostic checking of our results, and demonstrate that the models fitted by our approach yield residuals that are very close to Gaussian white noise, and thus have properties appropriate for prediction and control purposes.

2. RBF NETWORK-BASED MODELS
The RBF-ARX model may be written (Peng et al., 2001, 2003a)

\[
y_t = \phi_j(X_{t,j}) + \sum_{i=1}^{p_j} \phi_{ji}(X_{t,j}) y_{t-i} + \sum_{i=1}^{m} \Phi_{ui}(X_{t,j}) u_{t-i} + \epsilon_t
\]

where \(p_j, p_u\) are the model orders, \(m\) is the center order, and \(n_i\) is the dimension of the state \(X_{t,j}\). \(Z_{t,j} = (k = 1, 2, \ldots, m; i = y, u)\) are the centers of the RBF networks; \(\lambda^c_i, \lambda^p_i (k = 1, 2, \ldots, m)\) are non-negative scaling parameters, and \(\phi_0, \phi_{j,i}, \Phi_{ui}\) are state dependent coefficient scalars and vectors which are composed of RBF networks. \(c^c_i (k = 0, 1, 2, \ldots, m)\), \(c^p_i (k = 0, 1, 2, \ldots, m)\), and \(c^u_i (k = 0, 1, 2, \ldots, m)\) are the scalar or vector constants; and \(\|\|\|\) denotes the vector 2-norm. In model (1), the state \(X_{t,j}\) could be the output or the input signal, or both, or any other appropriate signal in the system to be considered. If there are no external inputs to the system, so that the coefficients \(\Phi_{ui}\) are all zero, the model reduces in form to an RBF-AR model (Vesin, 1993; Shi et al., 1999).

The estimation problem for the RBF-ARX consists of selecting appropriate initial values for the orders, scaling parameters and centers of the model, then using nonlinear optimization to find the best estimates of the parameters of the chosen model, on the basis of maximum likelihood. Different model orders may be compared on the basis of AIC, SBC and the results of diagnostic tests. It is especially important from the point of view of both prediction and control that the residuals after model fitting should be as close as possible to Gaussian white noise, and diagnostic tests should be applied to the residuals to check whether this is the case.

3. SELECTION OF ORDERS AND CENTERS
Although it is straightforward to estimate the parameters of the RBF-ARX model using nonlinear optimization, model (1) is very complex, and in practice the selection of the orders of the model, and of initial values for the centers could be very difficult. In this paper, we propose some simplifications of the RBF-ARX model, and a general approach for model identification.

3.1 The nonlinear behavior of \(y_t\) and its relation to the RBF-ARX coefficients
It may be shown (Ozaki and Oda, 1978; Haggan and Ozaki, 1981) that the nonlinear behavior of \(y_t\) depends on the movement of the characteristic roots of the model, that is by the movement of the roots of the equation

\[
z^{p_j} - \sum_{i=1}^{p_u} \Phi_{ui}(X_{t,j}) z^{-p_{u,i}} = 0
\]

Hence, the nonlinear behavior of \(y_t\) may be interpreted directly from \(\phi_{ji}\), whereas the effect of the nonlinear coefficients of the inputs, \(\Phi_{ui}\) cannot be directly interpreted, and may even interfere with the usual interpretation of equation (2). Furthermore, the inclusion of nonlinear \(\Phi_{ui}\) dramatically increases the number of parameters, and the consequent problem of choosing appropriate centers, with no clear interpretable benefit. Following the principle of parsimony, we suggest that it may be more appropriate to retain RBF-type coefficients of the form \(\phi_{ji}\) and \(\phi_{ui}\) but allow the coefficients \(\Phi_{ui}(X_{t,j})\) to be constant vectors, i.e. \(\Phi_{ui} = c^u_i\).

3.2 Extension of the linear approach to model order selection
Since model (1) is an extension of the linear ARX model, it follows that we could use the methods of linear estimation as an aid in order selection of the model. Thus, in the same way as for the linear case (Box and Jenkins, 1976), the autocorrelation, partial autocorrelation and cross-correlation functions of \(y_t\) may be useful as a guide to selecting the orders \(p_j\) and \(p_u\) of candidate models.

3.3 The interpretation of the centers
In many practical situations, the nonlinearity of \(y_t\) depends not only on the level of the state \(X_{t,j}\), but also on its rate of change, or velocity, with \(y_t\) exhibiting different behavior depending on whether the levels of the state are increasing or decreasing over time. In many applications such as those discussed in Peng et al. (2003a) the state dimension \(n_i = 2\), and the state

1211
$X_{t-1} = (x_{t-1}, x_{t-2})'$, where $x_t$, the megawatt load demand at time $t$, is one of the inputs of the RBF-ARX model. In order to examine the behavior of the process more explicitly as the level and velocity of $x_t$ changes, a modification of (1) for the case where $n_x = 2$, and $X_{t-1} = (x_{t-1}, x_{t-2})'$, would be to redefine the state so that $X_{t-1} = (x_{t-1}, \Delta x_{t-1})'$, where $\Delta x_t$ represents the velocity of $x_t$ and is given by $\Delta x_t = x_{t-1} - x_{t-2}$. Hence, the centers of this kind of RBF-ARX model should be expressed in terms of level and velocity.

If we go further and separate the two types of center, model (1) then becomes

$$
j_t = \phi(X_{t-1}) + \sum_{i=1}^{n_y} \phi_i(X_{t-1}) y_{t-1},
= \sum_{i=1}^{n_y} \phi_i(X_{t-1}) y_{t-1} + \epsilon_t
$$

$$
\phi(X_{t-1}) = \sum_{i=1}^{n_y} \phi_i(X_{t-1})
$$

$$
= c_0 + \sum_{i=1}^{n_y} \phi_i(X_{t-1})
$$

$$
\phi_i(X_{t-1}) = \sum_{k=1}^{n_y} \sum_{m=1}^{m} \phi_k(x_{t-1}) \exp[-(x_{t-1} - x_{k})^2]
$$

$$
= \sum_{i=1}^{n_y} \sum_{k=1}^{n_y} \phi_k(x_{t-1}) \exp[-(x_{t-1} - x_{k})^2]
$$

and $\lambda_{1}, x_{1}^{k}, \lambda_{2}, x_{2}^{k}, k = 1, 2, \ldots, n_x, \lambda_{3}, \Delta x_{3}^{k}, k = 1, 2, \ldots, n_t$ now represent scaling parameters and centers of level and velocity respectively.

From formulation (3), we can see that the effect of the coefficients, $c_k$, $c_k^{XX}$, $c_k^{Xx}$, $c_k^{XX}$ and consequently the effect of nonlinearity in the system, becomes smaller and smaller as $x_{t-1}$ and $\Delta x_{t-1}$ veer further and further away from the centers $x_{1}^{k}$ and $\Delta x_{2}^{k}$. Since the object of fitting the RBF-ARX model is to pick up on local variations and nonlinearities, one possible method of choosing the initial values of the centers is to use the modal values of the state levels and velocities. It also gives us a method for estimating how many centers are required, by observing how many modes (if any) the level and velocities in the state variable have. The modes themselves may be used as the initial values for the centers to be used in the nonlinear optimization procedure.

3.4 Diagnostic checking

After identifying the model orders and the number of centers of several candidate models, and finding initial estimates for the centers and scaling parameters (Shi et al., 1999), the models may now be estimated using nonlinear optimization. The most appropriate model may be selected by observing the residual sum of squares, the AIC and SBC values, and also by diagnostic checking. Since the objective of model estimation is prediction or control, which both require that residuals should be Gaussian white noise, the final model should be chosen after checking whether such an assumption is reasonable. The final selected model should be the one with a good overall performance in diagnostic tests as well as a small AIC. For series with a large sample size, the criterion SBC may also be a useful measure, since AIC tends to overfit the model in this situation (Schwarz, 1978).

4. APPLICATIONS

As an illustration of the practical application of the method, we will consider several applications. First of all, we will demonstrate our method by considering the fitting of an RBF-AR model to a discretization of the Mackey-Glass equation. We will also show how the method may be applied to examples of data taken from real thermal power plants, where NOx emissions are being modeled for the purpose of control.

4.1 RBF-AR Modeling of a Complex Nonlinear Time Series

The first example we shall consider is the discrete-time modeling of the Mackey-Glass equation

$$
y(t) = \frac{ay(t - \tau) - by(t)}{1 + y^2(t - \tau)} - dy(t)
$$

by an RBF-AR model of the form (3) where the state variable $X_{t-1} = (y_{t-1}, \Delta y_{t-1})'$.

The Mackey Glass equation is simulated for parameters $a = 0.2$, $b = 0.1$, $c = 10$ and $\tau = 20$, and discretized using the Euler method at time intervals $\Delta t = 0.1$. A thousand values are sampled at times $t = 1, 2, ..., 1000$ and the series is shown in Fig. 1. The first 500 data points are used to fit a model to the series, and the last 500 data points will be used to test the fit of the model.

In order to get an idea of the likely model order, $p_y$, we first examine the autocorrelation and partial autocorrelation function (Fig. 2) of the first 500 points of the simulated Mackey-Glass series. We can see that the autocorrelation function is dying away, whereas the major part of the partial autocorrelation function is cutting off at around six lags, although there are still some slightly significant values even up to lag 12.

The shape of the autocorrelation function indicates the possibility of seasonality, or high order autoregression. However, the partial autocorrelation function gives no support to the idea of seasonality, so high order autoregression seems to be more appropriate in this case. It therefore seems that a good approach to RBF-AR model-fitting for this series would be to look for $p_y$ in the range 6 – 12.

Now we have to consider initial estimates for the centers of the levels and differences of the Mackey-Glass series. In order to do this, we
construct histograms of the observations and differenced observations, which are shown in Fig. 3.

![Histograms](image1)

Fig. 1. A chaotic series generated from the Mackey-Glass equation

![Autocorrelation](image2)

Fig. 2. (top left) The autocorrelation function, b) (top right) partial autocorrelation functions (order 200), and (bottom) partial autocorrelation functions (order 25) of the Mackey-Glass

![Histogram](image3)

Fig. 3. Histogram of (left) the first 500 observations of the Mackey-Glass series, (right) the first 500 differences of the Mackey-Glass series

Observing the two histograms, we can see that there appear to be up to four modes in the histogram of the Mackey-Glass series itself, and up to four modes in the differenced Mackey-Glass series. However, after experimentation, it was found that good results could be obtained with only one mode on both cases. The initial value selected for the level center was 0.97, corresponding to the largest mode shown on the left of Fig. 3 and the initial value selected for the center for the differences was -0.035, corresponding to the largest mode shown on the right. The results of fitting RBF-AR models of different lag order with these centers are given in Table 1. First, models of lag order 6 through 12 were fitted. Since some residual autocorrelation remained after model fitting, higher order models (around orders 15 and 20) were also tried, and the best model from the point of view of small residual variance, small AIC and SBC, and uncorrelated residual errors was the model with lag order 19.

<table>
<thead>
<tr>
<th>$p_r$</th>
<th>Number of parameters</th>
<th>Residual Variance</th>
<th>AIC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>22</td>
<td>6.3899e-08</td>
<td>-8222.4</td>
<td>-7993.1</td>
</tr>
<tr>
<td>9</td>
<td>31</td>
<td>1.9810E-08</td>
<td>-8788.8</td>
<td>-8465.6</td>
</tr>
<tr>
<td>12</td>
<td>40</td>
<td>1.4107E-08</td>
<td>-8940.2</td>
<td>-8532.2</td>
</tr>
<tr>
<td>15</td>
<td>49</td>
<td>1.1401E-08</td>
<td>-9028.5</td>
<td>-8517.7</td>
</tr>
<tr>
<td>19</td>
<td>61</td>
<td>1.0182E-08</td>
<td>-9060.9</td>
<td>-8425.0</td>
</tr>
<tr>
<td>20</td>
<td>64</td>
<td>8.9974e-09</td>
<td>-9116.6</td>
<td>-8449.4</td>
</tr>
</tbody>
</table>

Table 1. RBF-AR models fitted to the Mackey-Glass series

![Residuals](image4)

Fig. 4. (top left) Residuals after fitting an RBF-AR(19,1,1) model to the Mackey-Glass series, (top right) the autocorrelation and partial autocorrelation function of the residuals, (bottom) the histogram of the residuals.

If we compare the results with those in Shi et al. (1999), and Peng et al. (2002) we can see that the residual error and AIC values for the fitted model are considerably smaller than those reported in those papers, although it is difficult to compare directly, since the simulations may be slightly different. Our best model has a somewhat larger number of parameters than most of the models suggested in those papers, but if we compare our models in general with those reported in Shi et al. (1999), and Peng et al. (2002) we find the residual variance is much reduced. The residuals after model-fitting are shown in Fig. 4, together with their autocorrelation and partial autocorrelation functions and histogram. The appearance of the residuals, and the shape of the residual autocorrelation function for our fitted model is very similar to that of white noise, although there are spikes in the residuals, and the histogram is rather heavy-tailed, whereas the residuals reported in Shi et al. (1999), and Peng et al. (2002) have a much more correlated appearance, and the autocorrelation function has not been reported. The prediction performance of our model on the test data set (the last 500 observations) was also good, with a residual prediction variance of 1.2306 x 10^-8.

4.2 Modeling of the Nitrogen Oxide (NOx) SCR Process in Thermal Power Plants

One example of a state dependent nonlinear
The Selective Catalytic Reduction (SCR) process in thermal power plants described by Peng et al. (2003a, 2003b, 2003c). When thermal power plants are used to produce electricity, nitrogen oxide NOx is emitted into the atmosphere, damaging the environment. There are strict regulations on how much NOx may be emitted, and in order to meet those regulations, the SCR process is usually controlled by means of a PID controller (see Fig. 5).

The device shown in Fig. 5 works by sensing the level of NOx ($v_{ij}$) coming directly from the turbine. This is an input to the PID controller which adjusts the amount of injected ammonia NH3 ($v_{ij}$), reducing the level of NOx actually output into the atmosphere, damaging the environment. There are strict regulations on how much NOx may be emitted, and in order to meet those regulations, the SCR process is usually controlled by means of a PID controller (see Fig. 5).

As suggested in section 2, we could use an RBF-ARX model of the form (3) to model this situation, which in terms of the process variables may be written

$$ y_t = \phi(X_{t,1}) + \sum_{i=1}^{m_p} \phi_i(X_{t,1})v_{t,i} $$

$$ + \sum_{k=1}^{m_w} \epsilon_{t,k} $$

where $u_t = (w_t, v_{1t}, v_{2t})'$ and the state $X_{t,1} = (w_{t-1}, \Delta w_{t-1})'$ where $w_t$ is the megawatt load demand at time $t$. As mentioned earlier, one of the major problems in fitting this kind of model is selecting the model orders, $p_w$, $p_v$, and the scaling parameters and centers

$$ \lambda_k^w, \omega_k^w; k = 1, \ldots, m_p, \lambda_k^v, \Delta \omega_k^v; k = 1, \ldots, m_v. $$

As in the previous examples, we obtain initial estimates of the lag orders by observing the autocorrelations, partial autocorrelations, and cross-correlations between the processes. These indicate that lag orders $p_w = p_v = 6$ are reasonable for this process. The histograms of the level and velocities of the megawatt load demand (Fig. 7) indicate three centers (initial values 78, 92, 115) are appropriate to represent the nonlinearities associated with megawatt level, and one center (zero) to represent the nonlinearities associated with velocities.

As mentioned earlier, one of the major problems in fitting this kind of model is selecting the model orders, $p_w$, $p_v$, and the scaling parameters and centers

$$ \lambda_k^w, \omega_k^w; k = 1, \ldots, m_p, \lambda_k^v, \Delta \omega_k^v; k = 1, \ldots, m_v. $$

As in the previous examples, we obtain initial estimates of the lag orders by observing the autocorrelations, partial autocorrelations, and cross-correlations between the processes. These indicate that lag orders $p_w = p_v = 6$ are reasonable for this process. The histograms of the level and velocities of the megawatt load demand (Fig. 7) indicate three centers (initial values 78, 92, 115) are appropriate to represent the nonlinearities associated with megawatt level, and one center (zero) to represent the nonlinearities associated with velocities.

Fig. 5. Selective Catalytic Reduction System for NOx emissions in GT-ST Combined Cycle Power Plants

environment to a much lower level $y_t$. From Fig. 5, we can see that the PID controller acts only on the input NOX, $v_{ij}$. However, the level of output NOx emissions also exhibits nonlinear dynamics dependent on the megawatt load demand of the plant, $w_t$. Hence $y_t$ depends on $w_t$, $v_{1t}$, $v_{2t}$, and from Fig. 6 we can see that $w_t$, the megawatt load demand fluctuates wildly, so that constructing a linear relationship between the process variables is problematic.

As suggested in section 2, we could use an RBF-ARX model of the form (3) to model this situation, which in terms of the process variables may be written

$$ y_t = \phi(X_{t,1}) + \sum_{i=1}^{m_p} \phi_i(X_{t,1})v_{t,i} $$

$$ + \sum_{k=1}^{m_w} \epsilon_{t,k} $$

where $u_t = (w_t, v_{1t}, v_{2t})'$ and the state $X_{t,1} = (w_{t-1}, \Delta w_{t-1})'$ where $w_t$ is the megawatt load demand at time $t$. As mentioned earlier, one of the major problems in fitting this kind of model is selecting the model orders, $p_w$, $p_v$, and the scaling parameters and centers

$$ \lambda_k^w, \omega_k^w; k = 1, \ldots, m_p, \lambda_k^v, \Delta \omega_k^v; k = 1, \ldots, m_v. $$

As in the previous examples, we obtain initial estimates of the lag orders by observing the autocorrelations, partial autocorrelations, and cross-correlations between the processes. These indicate that lag orders $p_w = p_v = 6$ are reasonable for this process. The histograms of the level and velocities of the megawatt load demand (Fig. 7) indicate three centers (initial values 78, 92, 115) are appropriate to represent the nonlinearities associated with megawatt level, and one center (zero) to represent the nonlinearities associated with velocities.

As in the previous examples, we obtain initial estimates of the lag orders by observing the autocorrelations, partial autocorrelations, and cross-correlations between the processes. These indicate that lag orders $p_w = p_v = 6$ are reasonable for this process. The histograms of the level and velocities of the megawatt load demand (Fig. 7) indicate three centers (initial values 78, 92, 115) are appropriate to represent the nonlinearities associated with megawatt level, and one center (zero) to represent the nonlinearities associated with velocities.

Fig. 6. Observations on $y_t$, $w_t$, $v_{1t}$, $v_{2t}$ taken from a real thermal power plant.

Fig. 7. Histograms of $w_t$ (left) and $\Delta w_t$ (right).

Fig. 8. Plot of residuals and autocorrelation function of the residuals after fitting the RBF-ARX model.

Fig. 9. Histogram of residuals and autocorrelation function of the residuals after fitting the RBF-ARX model.

The residuals after model-fitting are close to white noise as may be seen from Figs. 8 and 9. Fig. 8 shows the residuals themselves, together with their autocorrelation function. It may be seen that the residuals have a random appearance with no particular pattern, and the
autocorrelation function is similar to that of white noise. The correlations may also be shown to be insignificant using the Portmanteau statistic (Box and Pierce, 1970). Fig. 9 shows the histogram of the residual, which has a Gaussian appearance. In the thermal power plant case, the residuals not only appear to be white noise, they may be also shown to satisfy Normality (Jarque and Bera, 1987) and ARCH tests (Engle, 1982) as well as having insignificant autocorrelations. In fact the RBF-ARX model was developed for this kind of case, common in practice, where the relationship between variables is basically linear, but nonlinearities are introduced because of the erratic behaviour of one of the processes. We can therefore be satisfied that the fitted RBF-ARX model will perform well both for prediction and control, and in fact the model resulting from this procedure has been used in nonlinear model-based predictive control of the NOx emissions of a thermal power plant with excellent results (Haggan-Ozaki et al., 2005).

5. CONCLUSIONS
We have shown that the proposed systematic approach to fitting the RBF-ARX model works well in practice. The results reported compare well in terms of performance with the models reported by previous researchers (Shi et al., 1999; Peng et al., 2002). Moreover, we have outlined a step-by-step method which could be followed with some success by practitioners who are not experts in the field. By emphasizing diagnostic checking as an important and essential aid to model-fitting, we hope to help researchers find parsimonious models having the properties which are essential for successful prediction and control.

ACKNOWLEDGEMENT
The authors are grateful to Hideo Shioya and his colleagues at Bailey-Japan Co. Ltd. for providing the real thermal power plant data used in this paper.

REFERENCES