Chapter 5
Case Study 1 - Salinity Forecasting in the River Murray

5.1 Introduction

In this chapter, the state-of-the-art deterministic and Bayesian ANN development approaches are each applied to develop a model for forecasting salinity concentrations in the River Murray at Murray Bridge, South Australia, 14 days in advance. Considerable research has previously been conducted using this case study (*Maier*, 1995; *Maier and Dandy*, 1996, 1998a,b, 1999, 2000b; *Bowden*, 2003; *Bowden et al.*, 2002, 2003, 2005b); thus, it was considered to provide a good benchmark against which the methods proposed in this research could be evaluated and compared. Furthermore, salinity is a major water quality issue in Australia where there is a high dependence on river systems for consumptive uses; yet, significant negative impacts result due to the high salinity levels in these rivers. Having a reliable model for forecasting salinity concentrations is an integral part of this natural resources management issue.

5.2 Background

The Murray-Darling Basin (MDB), shown in Figure 5.1, is Australia's largest river system, covering an area of over one million square kilometres and extending into Queensland, New South Wales, Victoria, South Australia and the Australian Capital Territory. It supplies water for agriculture, domestic, industrial and environmental purposes and accounts for approximately 75% of all irrigation water use in Australia (see Figure 5.1 for irrigated areas within the MDB) (*Crabb*, 1997). The River Murray is the second longest river in the Basin and is South Australia's major surface water resource, providing 52% of the state's total annual water requirements on average (*Crabb*, 1997). In its natural state,

NOTE: This figure is included on page 216 of the print copy of the thesis held in the University of Adelaide Library.

Figure 5.1 Murray-Darling Basin (Source: Adapted from *MDBMC* (1999)).

the flow regime of the River Murray is highly variable, which is typical of large semiarid lowland river systems. However, it is now regulated by a series of storages, locks and weirs in order to maintain relatively constant pool levels, which enables pumping for irrigation and water supply, as well as navigation. As shown in Figure 5.2, water is diverted from the River Murray and supplied to many major South Australian towns, including the capital city of Adelaide, via five pipelines, with offtakes at Morgan, Swan Reach, Mannum, Murray Bridge and Tailem Bend.

5.2.1 Salinity in the River Murray

Salinity is a natural feature of the MDB. The groundwaters in the region through which the River Murray passes are generally highly saline, with levels not dissimilar to that of sea water (*EWS*, 1978). However, as a result of human activities and land use changes within the MDB since European settlement, the salinity levels in the River Murray have

NOTE: This figure is included on page 217 of the print copy of the thesis held in the University of Adelaide Library.

Figure 5.2 Pipelines delivering River Murray water to South Australia (Source: Adapted from *Crabb* (1997)).

been increasing, such that salinity is now a major water quality issue and a growing concern for users of the river's water. Among the causes of the increased salinisation of the River Murray are the regulation of the river and changes to the natural flow regime, the increased diversion of water for consumptive uses, unsustainable irrigation practices and clearance of vegetation within a large proportion of the catchment. Drainage flows from irrigation areas and rising groundwater levels due to dryland farming have increased the natural saline groundwater accessions, particularly in the South Australian section of the river (*Crabb*, 1997). Under natural conditions, this saline water would be flushed out to sea; however, as a result of river regulation and the substantial diversions of water both within South Australia and in the upstream states, there is often a lack of flushing, which then leads to the accumulation of salt in the lower reaches of the river (*MDBMC*, 1999). Consequently, there is a general increase in salinity with distance downstream. This is

indicated in Figure 5.3, which shows the 1998 mean salinity levels in the River Murray, in relation to the distance upstream from the Murray Mouth, together with salinity predictions for the years 2020, 2050 and 2100. These predictions were made as part of the Basin Salinity Audit of current and future threats of salinity to the MDB based on the salinity management strategies that were in place in 1998 (*MDBMC*, 1999). The salinity concentration in water is most widely estimated by electrical conductivity, where one electrical conductivity unit, or EC unit, is defined as one micro-Siemen per centimetre, or 1μS/cm, at 25° Celcius.

As the salinity of the River Murray increases, the options for consumptive uses of the water reduce due to the negative impacts on domestic, industrial and agricultural users at different salinity levels. In most cases, water with salinity concentrations greater than 700 EC is unsuitable for irrigating most horticulture crops and can lead to reductions in crop yields (*Crabb*, 1997), while from 800 EC, it becomes increasingly difficult to manage irrigation. A salinity concentration of 800 EC is also the upper limit for desirable drinking water quality set by the World Health Organization (WHO) (*MDBMC*, 1999), while 830 EC is considered to be the accepted maximum level for domestic supplies in large towns and cities (*Crabb*, 1997). For some small communities there is often no alternative water supply and although water with salinity concentrations greater than 800 EC may be safely consumed for short periods of time, the taste is objectionable (*MDBMC*, 1999). Other negative impacts due to high salinities include adverse biological effects, corrosion of pipes and infrastructure and increased consumption of soap and detergents due to the increased hardness of the water.

NOTE: This figure is included on page 218 of the print copy of the thesis held in the University of Adelaide Library.

Figure 5.3 Increasing salinity in the River Murray with distance downstream (Source:Adapted from *MDBMC* (1999)).

To address the salinity problem in the River Murray, in 1989 the Murray-Darling Basin Ministerial Council (MDBMC) adopted a Salinity and Drainage Strategy for the MDB. The aim of the Strategy was to provide a framework for reducing salinities in the river through cost effective salt interception and drainage diversion schemes (*MDBC*, 1999). Morgan, located some 320 km upstream from the Murray Mouth in South Australia, was selected as an indicator site to measure the performance of the Strategy, since Morgan is downstream of all major salt intakes and is the downstream limit of major horticultural and viticultural irrigation areas (*EWS*, 1978). The objective of the Strategy, which has been extended to 2015 (*MDBC*, 2001), is to maintain salinities less than 800 EC at Morgan for 95% of the time, as a 1 EC unit increase in Morgan salinity can result in significant economic impacts on downstream users. While, the impacts on various users of the River Murray water can vary depending on the locations of salt discharge, it has been estimated that more than 90% of both agricultural and non-agricultural impacts are borne by users in South Australia (*MDBC*, 1999). Figure 5.4 shows the estimated economic impacts incurred by river water users from various river reaches for a 1 EC unit increase at Morgan due to a salt discharge near Swan Hill in Victoria.

NOTE: This figure is included on page 219 of the print copy of the thesis held in the University of Adelaide Library.

Figure 5.4 Economic impacts borne by users of River Murray water for a 1 EC unit increase in Morgan salinity due to salt discharge near Swan Hill, Victoria (Source: *MDBC* (1999)).

5.2.2 Forecasting Salinity in the River Murray with ANNs

In the catchment water management plan for the River Murray in South Australia it is stated that "addressing dryland and irrigation-induced salinity is the most obvious and pressing problem that requires a significant, sustained commitment from the governments and community within the Basin" (*RMCWMB*, 2003). While long term management strategies take effect, an immediate option for reducing the economic impacts of high salinities on South Australian users is to make short term adjustments to pumping policies, allowing more water to be supplied in times of low salinity concentrations and less when salinity concentrations are high (*Dandy and Crawley*, 1992). However, this requires an accurate salinity forecasting model.

With this in mind, *Maier* (1995) and *Maier and Dandy* (1996, 1998a) developed the first ANN models for forecasting salinity concentrations at Murray Bridge, 14 days in advance. Murray Bridge was chosen as the site at which salinity forecasts were obtained, since the offtake from the River Murray at Murray Bridge (see Figure 5.2) provides a significant proportion of Adelaide's water supply. Furthermore, as seen in Figure 5.4, the economic costs to both agricultural and non-agricultural users due to salinity are high when water is extracted from the river at Murray Bridge. The forecasting period of 14 days was selected, as this is the minimum time required to make short term adjustments to the pumping schedule (*Maier and Dandy*, 1996). Of the models developed in these original studies, it was found that a one hidden layer ANN containing 51 inputs and 30 hidden nodes, trained using backpropagation, had the best generalisability. The original input set, consisting of 141 inputs, was selected based on *a priori* knowledge and then reduced using the results of a sensitivity analysis (*Maier and Dandy*, 1996), while the "optimum" number of hidden nodes was determined by trial and error, using the out-of-sample RMSE to evaluate generalisability (*Maier and Dandy*, 1998a). The minimum number of hidden nodes considered in these studies was 15.

Bowden et al. (2002) extended the salinity data set used by *Maier* (1995) and *Maier and Dandy* (1996, 1998a), such that it was approximately twice the length of the original, as shown in Figure 5.5. However, rather than increasing the sizes of the data subsets used to train, test and validate the ANN model developed, only the original data were used for model development, while the new data were reserved for simulating a real-time forecasting situation using the developed ANN model. In each of the studies conducted by *Bowden* (2003) and *Bowden et al.* (2002, 2003, 2005b), greater efforts were made at optimising the model development process and the new salinity data were used to assess the generalisability of the ANN models developed, given the different approaches inves-

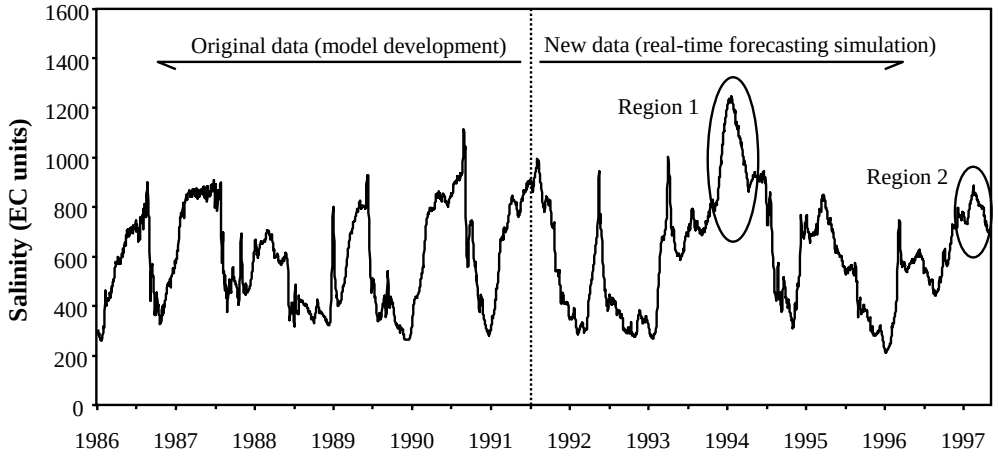

Figure 5.5 Extended time series of salinity in the River Murray at Murray Bridge, showing uncharacteristic data highlighted by regions 1 and 2.

tigated. It was found, in all of these studies, that each of the models developed performed poorly in the same two regions when applied to the new data, regardless of the model development methods used. These regions are highlighted by regions 1 and 2 in Figure 5.5. To diagnose the models' poor predictive performance in these regions, *Bowden et al.* (2002) clustered the entire data set (including both the original and new data) into groups of similar input/output patterns using a SOM (see Section 3.2.2). It was discovered that data from region 1 were contained in six clusters, all of which only contained new data (i.e. no data representative of region 1 had been used to develop the ANN models). Similarly, data from region 2 were also contained in six clusters, five of which only contained new data. Therefore, it was concluded that data contained regions 1 and 2 were uncharacteristic of the data used to develop the models and, consequently, the ANN models were required to extrapolate during these periods, resulting poor predictive performance. By inspecting the Fourier series for the seasonally varying mean salinity, developed using the training data, *Bowden et al.* (2003) identified that fluctuations in this series did not follow the seasonality of the new salinity data, as shown in Figure 5.6 (a), where the seasonal mean is moving out of phase with the actual salinity time series, particularly in regions 1 and 2. It was also found that the uncharacteristic high salinity concentrations in regions 1 and 2 corresponded to uncharacteristic low flow events in the new data, as seen in Figure 5.6 (b), which compares the Fourier series for the seasonally varying mean flow to the actual flow time series at Overland Corner, South Australia. This helps to explain why the data in regions 1 and 2 are uncharacteristic of the model development data.

NOTE: This figure is included on page 222 of the print copy of the thesis held in the University of Adelaide Library.

Figure 5.6 Fourier series seasonal mean versus actual time series for (a) salinity at Murray Bridge and (b) flow at Overland Corner (Source: Adapted from *Bowden et al.* (2003)).

In this research, an ANN model was developed for providing 14-day forecasts of salinity at Murray Bridge, using both deterministic and Bayesian methods. This case study was considered to be ideal for assessing the uncertainty associated with ANN predictions in a real-time simulation situation, and investigating the advantages of the proposed Bayesian training and model selection framework in comparison with state-of-the-art deterministic methods, particularly when the model is required to extrapolate. While an ANN based on a single weight vector may perform well in an interpolative context (on data similar to those contained in the training data set), it cannot be expected to extrapolate well in situations dissimilar to those previously presented to the model. However, it was hypothesised that, by accounting for the entire range of plausible weight vectors (those that provide a good fit to the data) using Bayesian techniques, a more generalised mapping to the underlying system function should be obtained, thus improving the extrapolation capabilities of the model developed.

5.3 AVAILABLE DATA AND MODEL INPUTS

Salinity in the River Murray at Murray Bridge is the sum of all salinity transported from upstream locations and saline accessions along the length of the river (*Maier and Dandy*, 1996). Salinity transport depends on flow rates and upstream salinity levels, while saline accessions, which are primarily due to inflows of saline groundwater, are a function of river level, groundwater level and flow. With the exception of groundwater levels, each of these variables are measured on a daily basis at various locations in the River Murray. *Maier and Dandy* (1996) considered that, because groundwater levels change very slowly, they may be considered as constant and, therefore, unnecessary for modelling salinity. Consequently, the available data set for this case study comprised daily salinity, flow and river level data from various locations in the lower River Murray between 1 December 1986 to 1 April 1998. For the period 1 December 1986 to 30 June 1992, this data set was compiled by *Maier* (1995) and supplied by the Engineering and Water Supply Department of South Australia. The data set was then extended for the period 1 July 1992 to 1 April 1998 by *Bowden* (2003) using data supplied by the Murray-Darling Basin Commission and the South Australian Department for Water Resources. The available data are summarised in Table 5.1 together with the abbreviations used for each variable, which are consistent with those used by *Bowden* (2003); *Bowden et al.* (2003) and *Bowden et al.* (2005b).

The inputs used in this research were those also used in the salinity forecasting model developed by *Bowden et al.* (2005b). These inputs were selected using the PMI selection

Table 5.1 Daily data available for forecasting salinity in the River Murray at Murray Bridge.

Location	Data Type	Abbreviation
Murray Bridge	Salinity	MBS
Mannum	Salinity	MAS
Morgan	Salinity	MOS
Waikerie	Salinity	WAS
Loxton	Salinity	LOS
Lock 1 Lower	Flow	L1LF
Overland Corner	Flow	OCF
Lock 7 Lower	Flow	L7F
Murray Bridge	River Level	MBL
Mannum	River Level	MAL
Lock 1 Lower	River Level	L1LL
Lock 1 Upper	River Level	L1UL
Morgan	River Level	MOL
Waikerie	River Level	WAL
Overland Corner	River Level	OCL
Loxton	River Level	LOL

method described in Section 3.2.4, following a two step procedure. In the first step, the PMI algorithm was run, in turn, for each of the 16 candidate input variables given in Table 5.1, using time lagged values of the current variable as potential inputs. The maximum lag that was considered for each variable was 60 days; thus, the potential input set for variable $\mathbf{x}_k$ included the past values $(\mathbf{x}_{k,t-1}, \ldots, \mathbf{x}_{k,t-60})$. In this bivariate stage, PMI values were calculated between lagged values of each input variable and the output variable MBS_{t+13}, in order to determine the significant lags for each variable. The second step was a multivariate stage, where the significant lags for each variable selected in the first step were combined to form a new set of potential inputs. The PMI algorithm was then run again to determine the final set of important predictors. These inputs are given in Table 5.2, together with their PMI score, the RI values estimated based on the PMI scores and the corresponding rank order of importance. As it can be seen, 13 inputs were selected from the total set of 960 potential inputs (16 variables $\times$ 60 lags = 960).

As seen in Table 5.2, the salinity concentration at Mannum with a lag of 1 day (MAS_{t-1}) was found to be the most important input for forecasting salinity concentrations at Murray Bridge, 14 days in advance (MBS_{t+13}). This was followed by salinity at Waikerie, also with a lag of 1 day (WAS_{t-1}). During times of low flow (6500 ML/day), the travel time (time taken for salt to travel from one location to another) between Mannum and Murray Bridge is approximately 14 days, while, during times of intermediate flow (16,000–17,000 ML/day), the travel time between Waikerie and Murray Bridge is 14–16 days (*Maier and Dandy*, 1996). Therefore, it is not surprising that these two inputs

Table 5.2 Inputs used in salinity forecasting ANN model, together with the PMI scores, PMI-based RI estimates and rank order of importance.

Variable	Lag (days)	PMI Score	RI (%)	Rank Importance
MAS	1	1.269	70.32	1
MOS	60	0.326	1.20	9
WAS	1	1.188	14.54	2
WAS	43	0.410	1.51	6
LOS	25	0.401	1.89	5
L7F	1	0.747	2.44	3
MBL	1	0.264	1.41	7
MBL	11	0.307	0.93	10
MBL	21	0.280	0.76	13
MBL	34	0.265	0.79	12
MBL	57	0.277	0.83	11
MAL	57	0.310	1.33	8
L1UL	1	0.357	2.05	4

were found to be the most important, as it would be expected that the most significant inputs would have travel times between their upstream locations and Murray Bridge of approximately 14 days. The next most significant inputs were flow at lock 7 (which is located just upstream of the border between South Australia and Victoria) with a lag of 1 day ($\mathrm{L7F}_{t-1}$) and river level at lock 1 (located at Blanchetown) with a lag of 1 day ($\mathrm{L1UL}_{t-1}$), which are significant variables involved in determining groundwater accessions and salinity travel times. Although also found to be significant by the PMI algorithm, lags of river level at Murray Bridge were found to be the least important predictors, with the exception of a lag of 1 day.

Time series plots of the significant inputs versus salinity at Murray Bridge at time $t + 13$ (MBS_{t+13}) between 1987 and 1992 are shown in Figures 5.7 to 5.14. Where more than 1 lag of an input variable was found to be significant (e.g. WAS, MBL), only plots of the most significant lag, according to PMI score, are shown. It can be seen in Figures 5.7 to 5.9 that salinity concentrations at Mannum, Morgan and Waikerie are generally similar (although slightly lower) to those at Murray Bridge, with the same seasonal fluctuations. It can also be seen in Figure 5.10 that salinity concentrations at Loxton are significantly lower than those at downstream locations, which indicates that large saline accessions occur between Loxton and Waikerie. This may, in part, be due to raised ground water levels resulting from the irrigation in this region (see Figure 5.1). In Figure 5.11, a strong inverse relationship between flow at lock

7 and salinity at Murray Bridge is evident. Figures 5.12 to 5.14 also display an inverse relationship between river level at various locations and salinity at Murray Bridge.

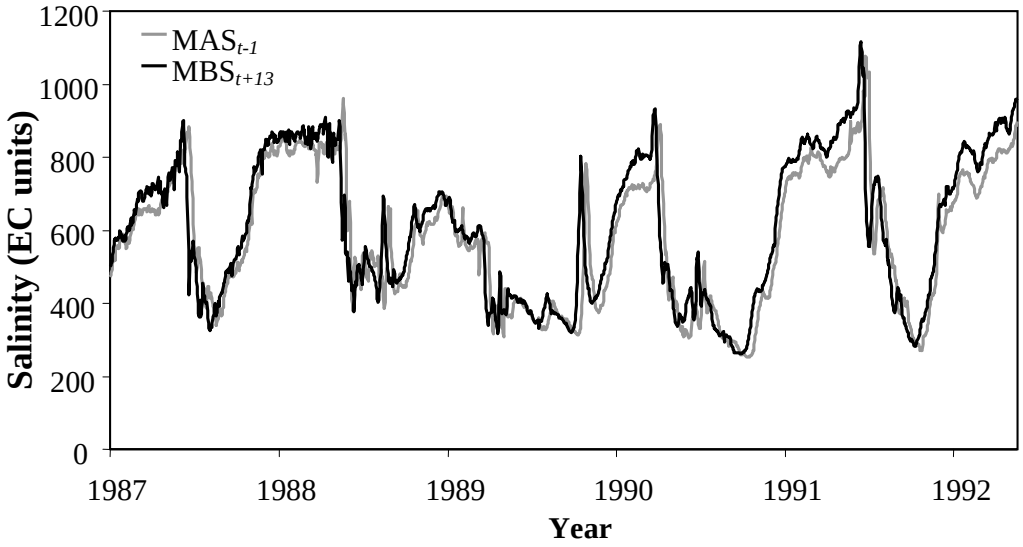

Figure 5.7 Salinity at Mannum with a lag of 1 day (MAS$_{t-1}$).

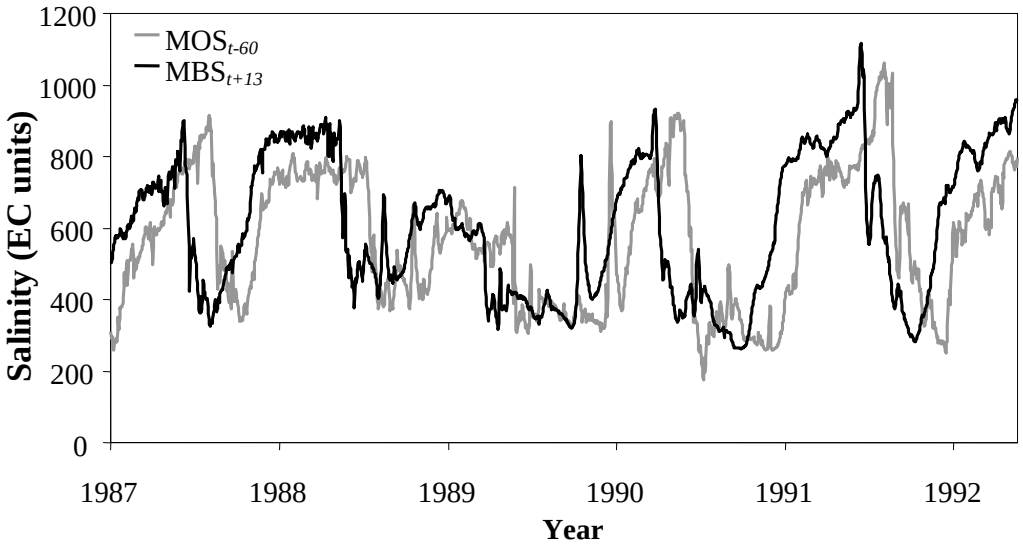

Figure 5.8 Salinity at Morgan with a lag of 60 days (MOS$_{t-60}$).

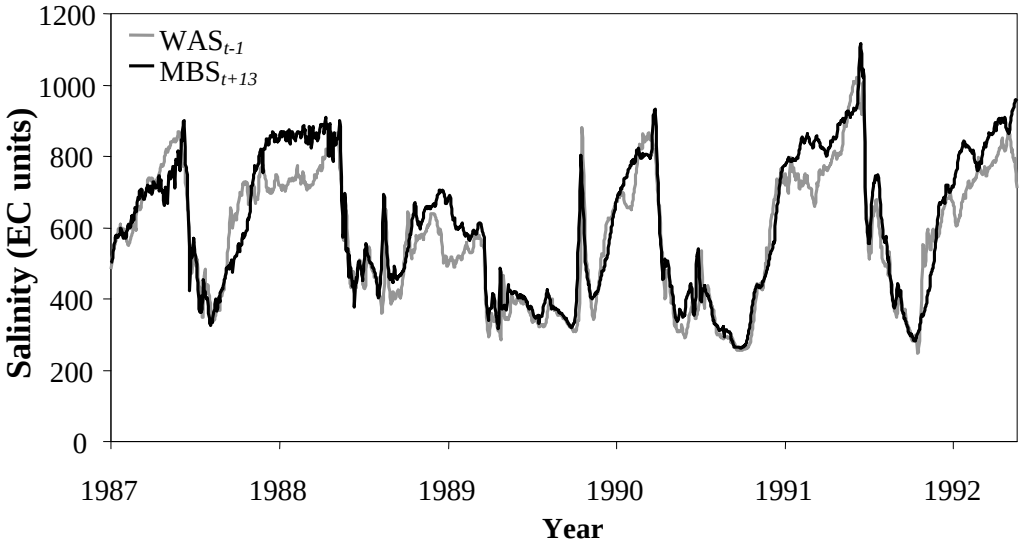

Figure 5.9 Salinity at Waikerie with a lag of 1 day (WAS$_{t-1}$).

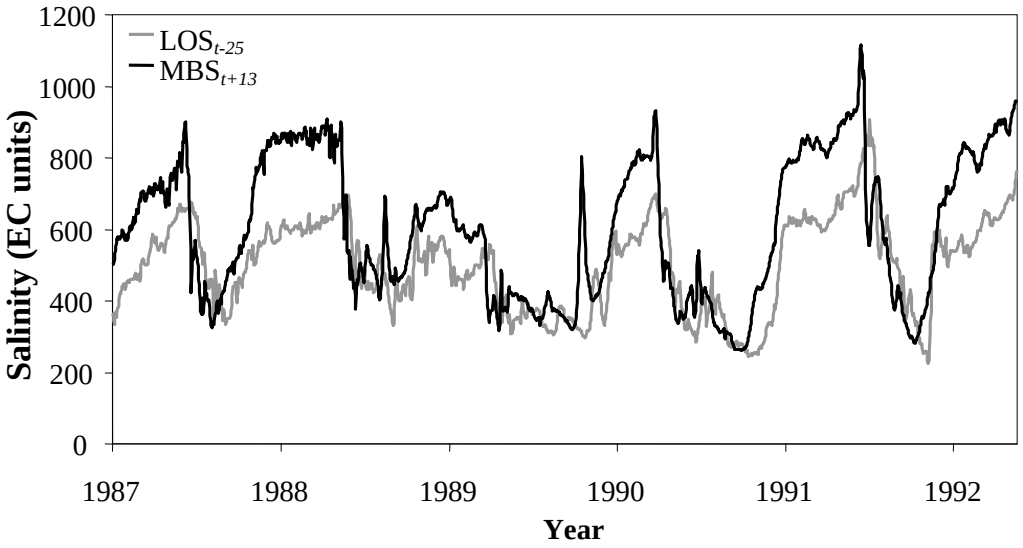

Figure 5.10 Salinity at Loxton with a lag of 25 days (MAS$_{t-1}$).

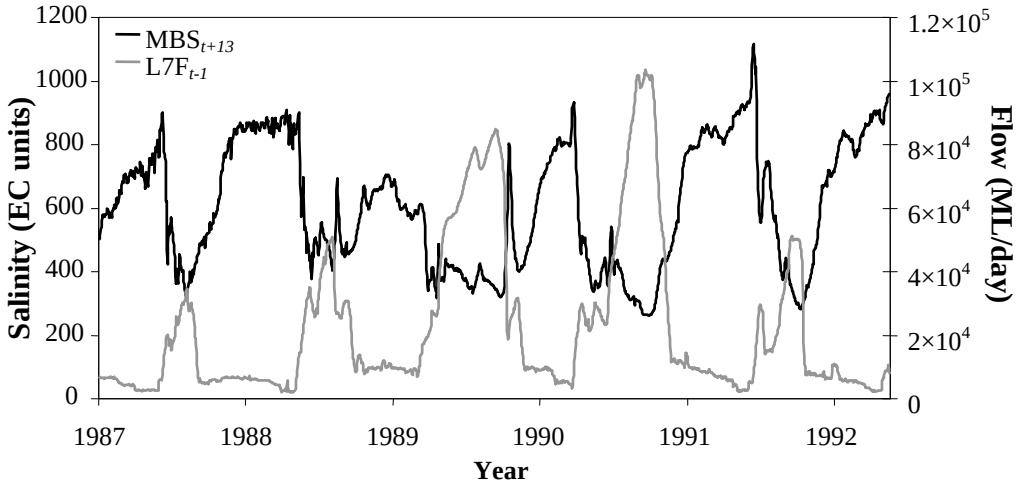

Figure 5.11 Flow at lock 7 with a lag of 1 day ($L7F_{t-1}$).

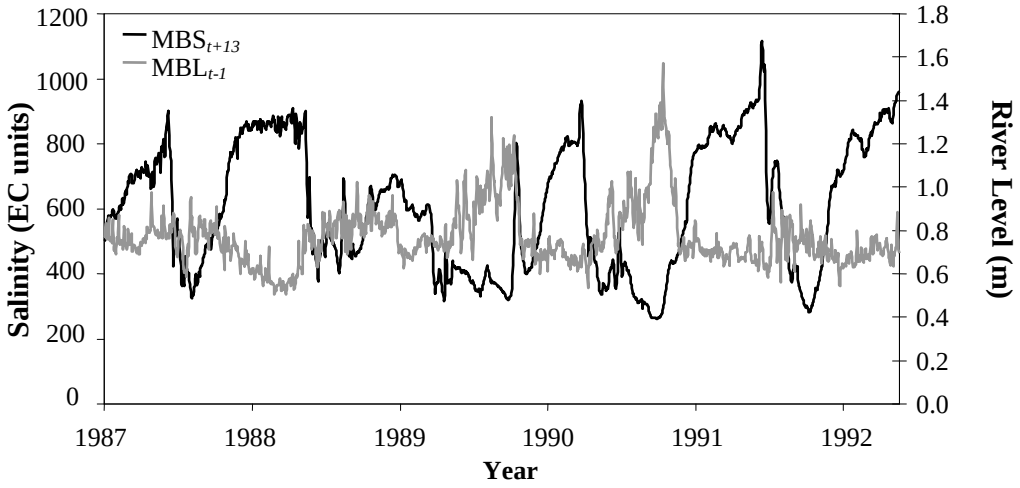

Figure 5.12 River level at Murray Bridge with a lag of 1 day (MBL_{t-1}).

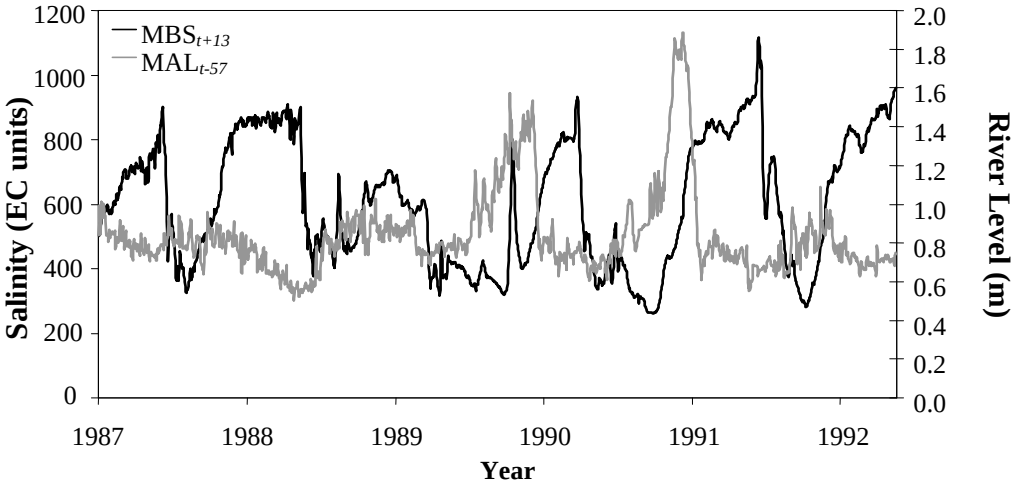

Figure 5.13 River level at Mannum with a lag of 57 days (MAL_{t-57}).

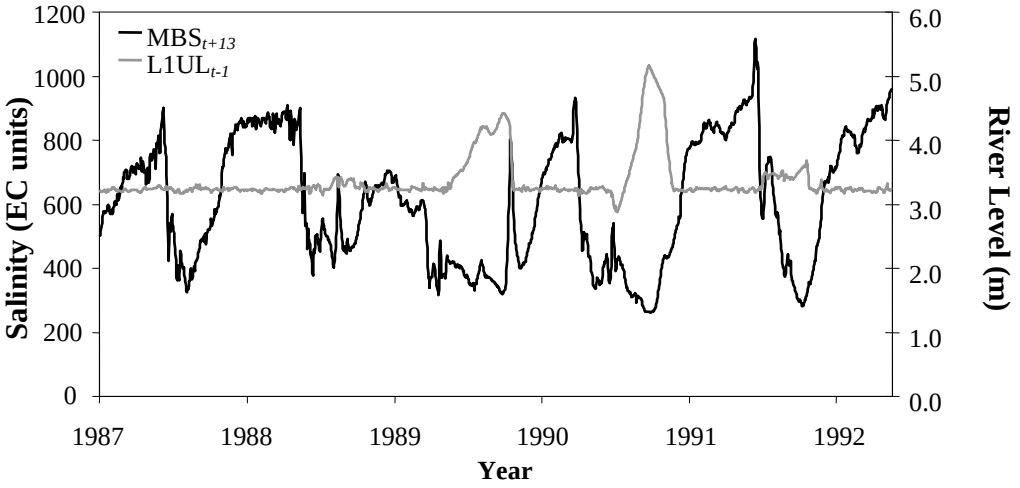

Figure 5.14 River level upstream of Lock 1 with a lag of 1 day (L1UL_{t-1}).

Similar to the studies by *Bowden* (2003) and *Bowden et al.* (2002, 2003, 2005b), in this research, data from 1 December 1986 to 30 June 1992 were used for model development, while data from 1 July 1992 to 1 April 1998 were reserved to simulate a real-time forecasting situation. After accounting for the appropriate lags of the input and output variables, there were 1964 data samples in the "model development" data and 2028 samples in the "real-time forecasting" data. The model development data were divided into training, testing and validation subsets using the SOM data division method discussed in Section 3.2.2. By comparing the average silhouette widths and discrepancy values (see Section 3.2.2.2) for various SOM grid sizes ranging from 1×2 to 12×12, it was found that a grid size of 8×12 was optimal for clustering this data set. This resulted in 87 clusters containing at least 3 samples, 2 clusters containing 2 samples and 1 cluster containing only 1 sample (6 grid cells were empty). From these, 1257 (64%) samples were allocated to the training data subset, 314 (16%) samples were allocated to the testing subset and the remaining 393 (20%) samples were allocated to the validation subset, ensuring that at least one sample from each cluster was allocated to each of the subsets where possible. As it can be seen in Figure 5.15, the salinity data were not normally distributed. However, initially, the only transformation applied to the input and output data was linear standardisation, such that each variable had a mean of zero and standard deviation of one. The appropriateness of this transformation was then checked once the preliminary models had been developed.

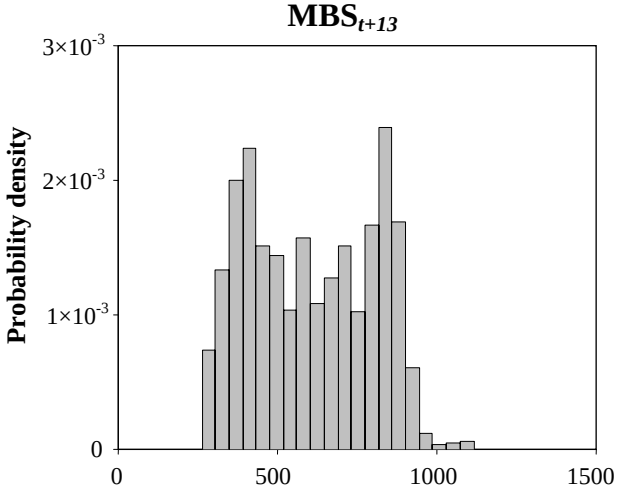

Figure 5.15 Histogram of Murray Bridge salinity data.

5.4 DETERMINISTIC ANN DEVELOPMENT

The state-of-the-art deterministic ANN methodology, described and developed in Chapter 3, was first applied to the salinity forecasting case study, taking into account the results obtained in the investigations carried out on the synthetic data sets.

5.4.1 Methods

5.4.1.1 Model Selection

Using the 13 inputs given in Table 5.2, *Bowden et al.* (2005b) found that a 32 hidden node ANN model containing 481 weights was optimal for producing 14-day salinity forecasts. In this research, the optimal ANN geometry was re-investigated, using the trial-and-error approach discussed in Chapter 3, in the attempt to obtain a more parsimonious model. The number of hidden nodes selected by *Bowden et al.* (2005b) was used to guide the trial-and-error search for the optimal geometry; however, since a more parsimonious model was sought in this research, only models containing less than 32 hidden nodes were considered. To minimise the number of networks that required training and testing, ANNs containing 5, 10, 15, 20, 25 and 30 hidden nodes were initially considered, in order to narrow down the search to a 10 hidden node range (e.g. 1–10, 11–20, or 21–30). Once this range was determined, the trial-and-error approach was repeated, with the number of hidden nodes increasing in increments of 2 (e.g. if the 10 hidden node range found was 1–10, the next trial would include networks of 2, 4, 6, and 8 hidden nodes). Finally, the search was reduced to single node increments, where the networks either side of best network found in the previous trial were tested (e.g. if the model with the best generalisability contained 8 hidden nodes, networks with 7 and 9 hidden nodes were also tested). As mentioned in Section 3.3, all of the ANN models developed in this research contained a single hidden layer with tanh hidden nodes and linear output nodes.

According to the findings of Section 3.4, the in-sample BIC and out-of-sample AIC were used to assess the generalisability of the models developed. However, in Section 3.4, it was inconclusive whether or not the in-sample BIC would adequately penalise model complexity in order to select the optimal model structure when applied to a complex, noisy real-world problem. This issue was further investigated in this study by comparing the in-sample BIC values obtained when training was run to convergence and when it was stopped early according to the test set error, in order to assess whether the degree of overfitting that occurred affected the size of the ANN that was identified as having the greatest generalisability. These values were also compared with the out-of-sample AIC values, to determine whether or not there was agreement between the different values,

resulting in the same optimal geometry being selected.

5.4.1.2 Training

The networks were trained with the SCE-UA algorithm, according to the results of the training algorithm comparison carried out in Section 3.4. However, due the number of inputs used in the salinity forecasting model, the dimension of the weight vector was relatively large, even for networks containing few hidden nodes (e.g. a 5 hidden node ANN contains 76 weights). As mentioned in Section 3.4.7, being a randomised search method, the SCE-UA algorithm is likely to be relatively inefficient when applied to high dimensional problems (a gradient-based search technique would generally be the first choice for such problems when the objective function is differentiable, even if it were less robust). It was recommended that, in order prevent excessively long training times, upper limits should be considered for the SCE-UA algorithm parameters when the dimension of the weight vector d is greater than 40. Therefore, the modifications given in Table 5.3 were applied in this study. It is possible that these modifications may have reduced the optimisation capabilities of the SCE-UA algorithm; however, due to time and computational requirements, training the large networks considered in this study (e.g. a 30 hidden node ANN contains 451 weights) would have been infeasible without these modifications. To increase the chances of obtaining a globally optimal solution, the algorithm was initialised with three different sets of random weights for each network trained. Furthermore, to verify that the results obtained using the modified SCE-UA parameters were reasonable, they were compared to those obtained by *Bowden et al.* (2005b) using a 32 hidden node ANN, trained by backpropagation.

Training was run until the stopping criterion given by (3.33) was met or after 10 mil-

Table 5.3 Modified parameter values for SCE-UA algorithm

Parameter	Value
m	$\begin{cases} 2d+1 & \text{if } d \leq 40 \\ 2 \times 40 + 1 & \text{otherwise} \end{cases}$
q	$\begin{cases} d+1 & \text{if } d \leq 40 \\ 40+1 & \text{otherwise} \end{cases}$
β	$\begin{cases} 2d+1 & \text{if } d \leq 40 \\ 2 \times 40 + 1 & \text{otherwise} \end{cases}$
p	$\begin{cases} d & \text{if } d \leq 40 \\ 40 & \text{otherwise} \end{cases}$

lion error function evaluations, whichever was first. Cross-validation using the test data set was also employed during training and the weights resulting in the minimum test set error were saved, and the corresponding model outputs computed.

5.4.1.3 Validation

The optimal model structure selected was validated by subjecting it to the independent validation data set that was not used for training or testing the model. The results obtained for the training, testing and validation subsets were then compared to those obtained by *Bowden et al.* (2005b). However, although the same proportions of data were allocated to the respective subsets, the data division method used in this research was different to that used by *Bowden et al.* (2005b); thus, different training, testing and validation subsets were produced. The selected model was also applied to the new "real-time forecasting" data, which not only allowed an assessment of how the model would perform in a real-time forecasting situation, but also enabled a comparison of the model's predictive performance to that of the 32 hidden node ANN model developed by *Bowden et al.* (2005b) on the same data set.

In order to assess how well the selected model had approximated the underlying input-to-output relationship, the modified Connection Weight Approach (see Section 3.4.4.3) was used to evaluate RI values for each of the model inputs, which were then compared to the PMI-based RI estimates given in Table 5.2.

5.4.2 Results

After the initial models containing $5, 10, \ldots, 30$ hidden nodes were trained, the standardised model residuals were inspected to determine whether or not a linear transformation of the input and output data was sufficient. Histograms of the standardised residuals for each of the networks developed are shown in Figure 5.16, in comparison to the standard normal distribution. It can be seen that, in each case, the standardised residuals are approximately normal with no significant outliers; thus, it was concluded that no further transformations of the data were necessary.

In this study, all models were developed using an Intel Xeon processor with 2 GB of RAM running at 2.4 GHz. The resulting average training times for the initial network sizes are given in Table 5.4. It can be seen that even with the upper limits placed on the algorithm parameters, the training times for the different networks were long. However, it was found that, for each training run, the algorithm was stopped according to the stopping criterion given in (3.33), rather than because the number of function evaluations had

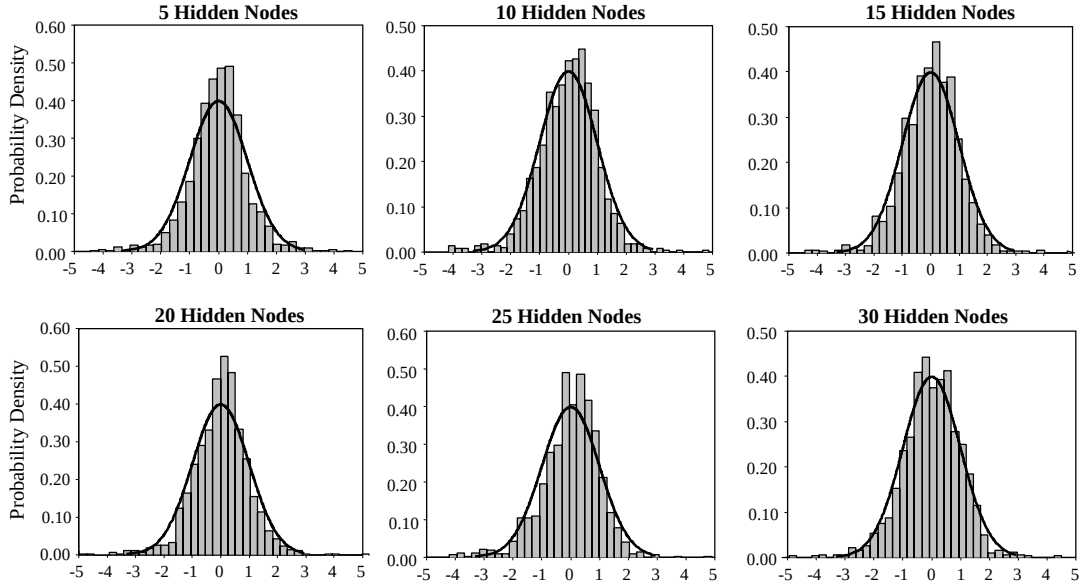

Figure 5.16 Standardised model residuals.

exceeded 10 million, as was often the case when large networks were trained with the SCE-UA algorithm without the modifications. This meant that training had always converged to a stationary point of the error surface. In comparison to the RMSE of 29.3 EC obtained by *Bowden et al.* (2005b) for the training data using a 32 hidden node ANN, all of the training data RMSE values obtained for the initial models were less than 29.0 EC, even though most of the models were significantly less complex than that used by *Bowden et al.* (2005b). Therefore, it was considered that the optimisation capabilities of the SCE-UA algorithm with parameter modifications had not been significantly diminished (although it should be kept in mind that the training sets were not exactly the same, as mentioned in Section 5.4.1.3).

Table 5.4 Training times using the SCE-UA algorithm.

Hidden Nodes	Training Time (hours)
5	3.279
10	4.939
15	7.613
20	9.597
25	11.123
30	17.416

The the out-of-sample AIC values and in-sample BIC values obtained when training was run to convergence, and when it was stopped early according to the test set error, are plotted against one another in Figure 5.17 for the different network sizes. It can be seen that the in-sample BIC values obtained when training was stopped early and when it was allowed to converge are quite different. The "converged" BIC values indicate that the models containing between 10 and 20 hidden nodes had the best generalisability, whereas the "stopped" BIC values indicate that this was achieved with the 5 hidden node ANN. It can also be seen that there is much better agreement between the out-of-sample AIC values and the in-sample BIC values when overtraining was prevented. It is also interesting to note that, for the 5 hidden node ANN, the "converged" and "stopped" in-sample BIC values were approximately the same, which suggests that any overfitting that occurred when training was run until convergence was relatively insignificant. Therefore, it was concluded that the 5 hidden node ANN was a more optimal structure for the salinity forecasting model than the larger network sizes. It was also concluded that, in order to obtain correct results using the in-sample BIC, prevention of overtraining is important; thus a test set is necessary when using this model selection criterion. Henceforth, the in-sample BIC will refer to the value calculated when training was stopped early.

Given these results, the trial-and-error search for the optimal model structure was

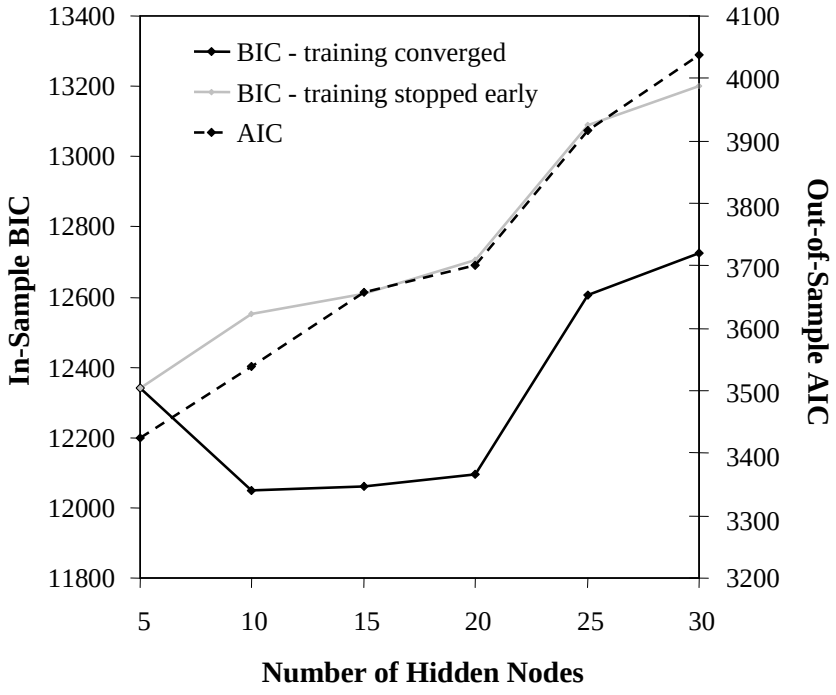

Figure 5.17 In-sample BIC and out-of-sample AIC values for $5, 10, \ldots, 30$ hidden node ANN models.

narrowed down to the 1–10 hidden node range and the subsequent ANN models trained and tested for their generalisability contained $2, 4, \ldots, 8$ hidden nodes. The resulting in-sample BIC and out-of-sample AIC values for these networks are plotted in Figure 5.18, where it can be seen that the 4 hidden node ANN was identified as having the best generalisability by each criterion. Finally, a 3 hidden node ANN was trained and tested, and the in-sample BIC and out-of-sample AIC results were compared for the 3, 4 and 5 hidden node ANNs. These values are plotted in Figure 5.19. It can be seen in this figure that, again, both model selection criteria identified the 4 hidden node as having the best generalisability. From these results, it was concluded that the 4 hidden node ANN was the optimal model structure for the salinity forecasting model and this model was therefore subjected to the independent validation set for verification. The 4 hidden node ANN contains 61 weights (i.e $d = 61$), which is significantly fewer than the 481 weights contained in the ANN model developed by *Bowden et al.* (2005b) for the same problem. Therefore, the aim of developing a more parsimonious model in this research was achieved. The optimal ANN structure was identified from a possible 30 different network sizes by only training and assessing 11 ANN models.

Shown in Figure 5.20, are the model predictions versus the (a) training data, (b) testing data and (c) validation data. It can be seen that the model provided a good fit to each

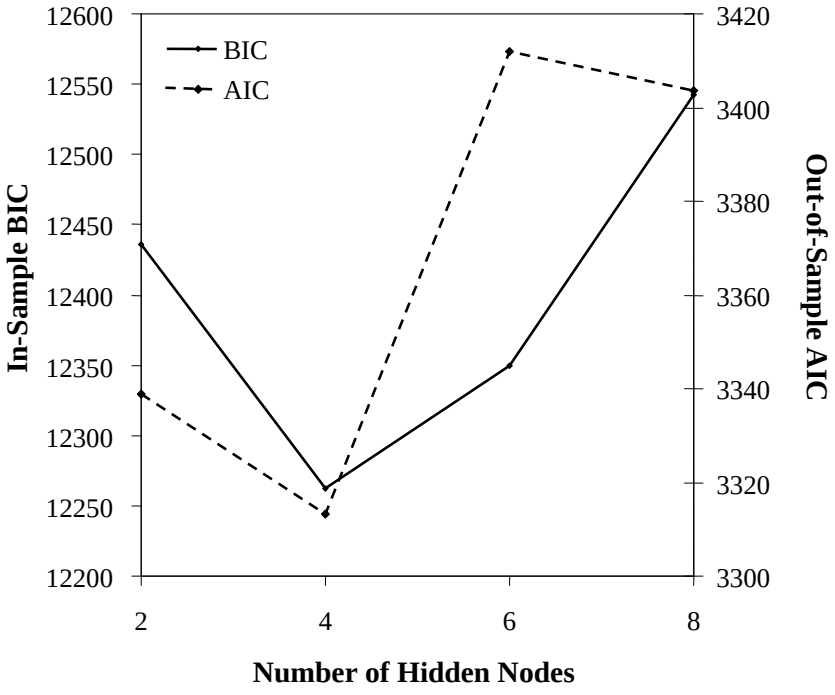

Figure 5.18 In-sample BIC and out-of-sample AIC values for 2, 4, 6 and 8 hidden node ANN models.

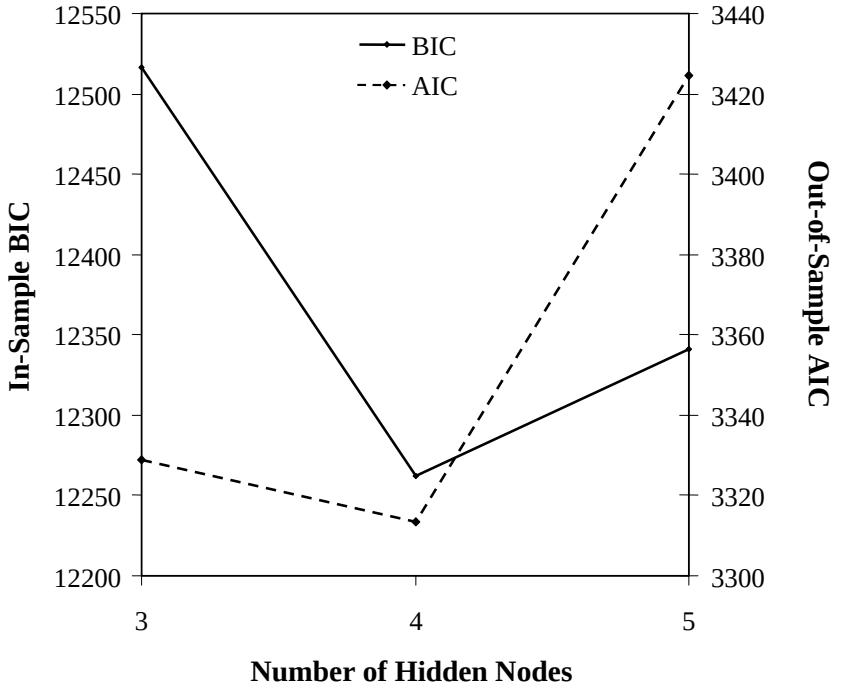

Figure 5.19 In-sample BIC and out-of-sample AIC values for 3, 4 and 5 hidden node ANN models.

of these data subsets; however, as seen in Figure 5.20 (a), salinity concentrations above 1000 EC units tended to be under-predicted by the model. Likewise, a number of points in the 400–600 EC range were under-predicted. It can be seen in Figure 5.21, which shows a time series plot of the 14-day forecasts against the recombined model development data

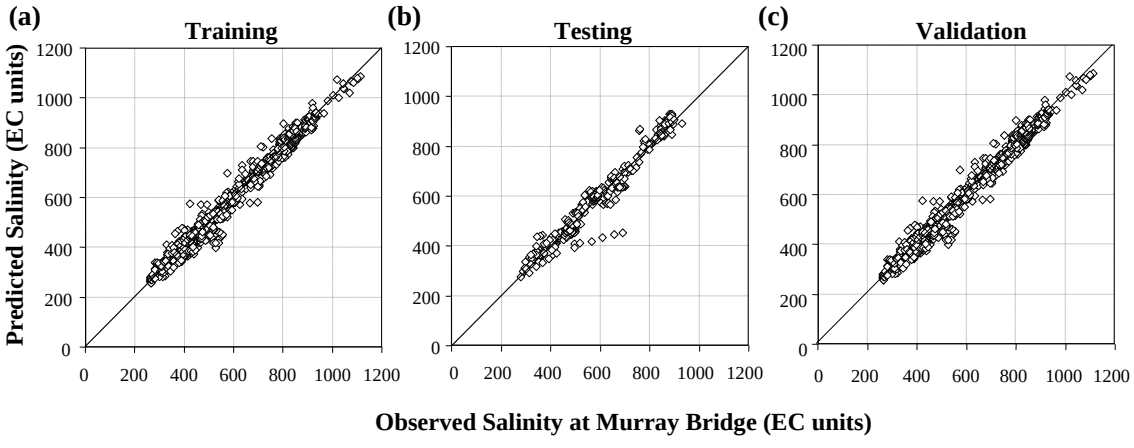

Figure 5.20 Scatter plots of the 4 hidden node ANN model predictions versus the (a) training, (b) testing and (c) validation data.

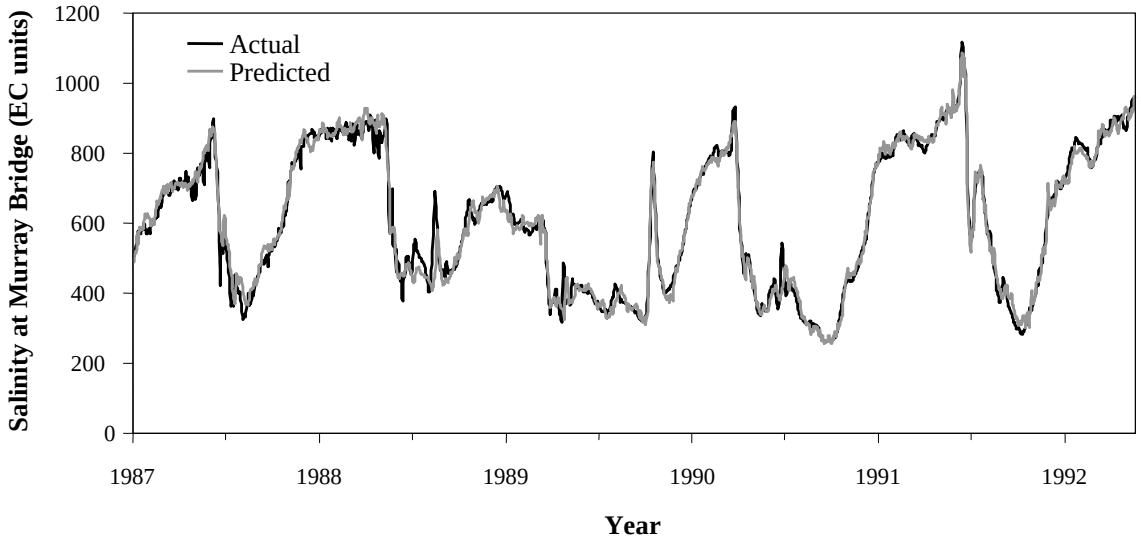

Figure 5.21 Time series plot of the 14-day salinity forecasts obtained with the 4 hidden node ANN model against the "model development" data.

(training, testing and validation), that within this range there are a number of small peaks in the observed salinity concentrations, which account for the under-predicted salinities. Nevertheless, it can also be seen in this plot that, overall, the observed salinity concentrations were predicted well by the model. The resulting MAE, RMSE and CE values for the training, testing and validation data sets are given in Table 5.8. The RMSE measure was used to assess model performance in *Bowden et al.* (2005b), thus its use in this research enabled a direct comparison of the results obtained using the 4 hidden node ANN model with those given in *Bowden et al.* (2005b) (also given in Table 5.8).

The model-based RI values for each of the model inputs, calculated using the modified Connection Weight Approach, are given in Table 5.5, together with the resulting rank order of modelled input importance. In comparison to PMI-based RI estimates (also given in Table 5.5), it can be seen that the inputs identified as being the most important by the model-free approach, also provided the greatest contributions to predicting the output using the ANN model. However, the model-based RI values are generally somewhat higher than the PMI-based RI estimates, indicating that the model has given greater importance to many of the inputs than the model-free estimation approach. It is considered that this is primarily due to the fact that the model inputs are not independent and that the model has shared the contributions of some inputs among a number of correlated variables. In the first step of the PMI algorithm, before any inputs have been selected, the mutual information (MI) between each of the inputs and the dependent out-

Table 5.5 Model-based RI estimates and order of input importance in comparison to the PMI- and MI-based RI estimates.

	Rank	*RI* (%)		
Input	Importance	Model-based	PMI-based	MI-based
MAS_{t-1}	3	13.01	70.32	17.99
MOS_{t-60}	11	3.90	1.20	6.36
WAS_{t-1}	1	17.23	14.54	16.33
WAS_{t-43}	10	3.91	1.51	9.64
LOS_{t-25}	4	8.86	1.89	11.30
$\mathrm{L7F}_{t-1}$	2	16.92	2.44	8.73
MBL_{t-1}	13	1.76	1.41	5.21
MBL_{t-11}	8	5.62	0.93	4.78
MBL_{t-21}	7	6.18	0.76	4.50
MBL_{t-34}	12	3.10	0.79	4.13
MBL_{t-57}	9	4.37	0.83	3.69
MAL_{t-57}	6	6.83	1.33	4.02
$\mathrm{L1UL}_{t-1}$	5	8.31	2.05	3.31

put is computed. The RI values calculated based on these MI values are also given in Table 5.5. Unlike the PMI-based RI estimates, these values measure the amount of information about the output contained in each of the inputs, regardless of whether this same information is also contained in any of the other inputs (i.e. they do not only measure the additional information contained in each input). It can be seen in Table 5.5 that the MI-based RI estimates are similar in magnitude to the modelled input contributions, which indicates that similar information from correlated inputs is indeed being used to model the output. It can also be seen that the MI between inputs MAS_{t-1} and WAS_{t-1} and the output MBS_{t+13} is very similar; yet, because MAS_{t-1} was selected as the most important input, the PMI criterion estimated that over 70% of the information needed to predict the output was provided by this input, while only an additional 14.5% was provided by WAS_{t-1}. If the MI between WAS_{t-1} and MBS_{t+13} had been slightly higher, it would have been given the larger contribution, while only additional information provided by MAS_{t-1} would have been considered. It is clear that relative input importance, as estimated by the PMI criterion, is very much dependent upon the order in which inputs are selected, and this is, to some extent, dependent upon the available data and the calculation of the PMI criterion. Nevertheless, as the model-based RI values were of approximately the same order as the PMI-based RI values, and approximately the same magnitude as the MI-based estimates, it was considered that the model had captured important physical information about the underlying system.

Being satisfied with the model validation results, the 4 hidden node ANN model was applied to the "real-time forecasting" data to assess how well the model would perform in a real-world application. A plot of the resulting 14-day forecasts obtained is shown in Figure 5.22. It can be seen that the most notable regions of poor predictive performance were regions 1 and 2, which were identified as being uncharacteristic by *Bowden et al.* (2002), but apart from these regions, the model was able to capture the variations in salinity concentration over the forecasting period. However, the fact that the model was not robust to the uncharacteristic data, means that poor management decisions could be made if this model was relied on in a real-world application. The WHO 800 EC threshold for desirable drinking water quality is also shown in this figure, as salinities above this level can have significant negative impacts, as discussed in Section 5.2.1. It can be seen that the actual salinity concentrations in regions 1 and 2 are greater than 800 EC units, while the 14-day forecasts indicate salinity levels less than this threshold. If pumping schedules were altered according to these forecasts in order to minimise the negative impacts due to high salinity, it is possible that these alterations may, in fact, have had the reverse effect, as more water would be pumped in times of high salinity concentrations. The major limitation of these deterministic forecasts is that there is no measure of confidence provided that indicates their quality, except for the fit to the model development data, which was seen to be very good.

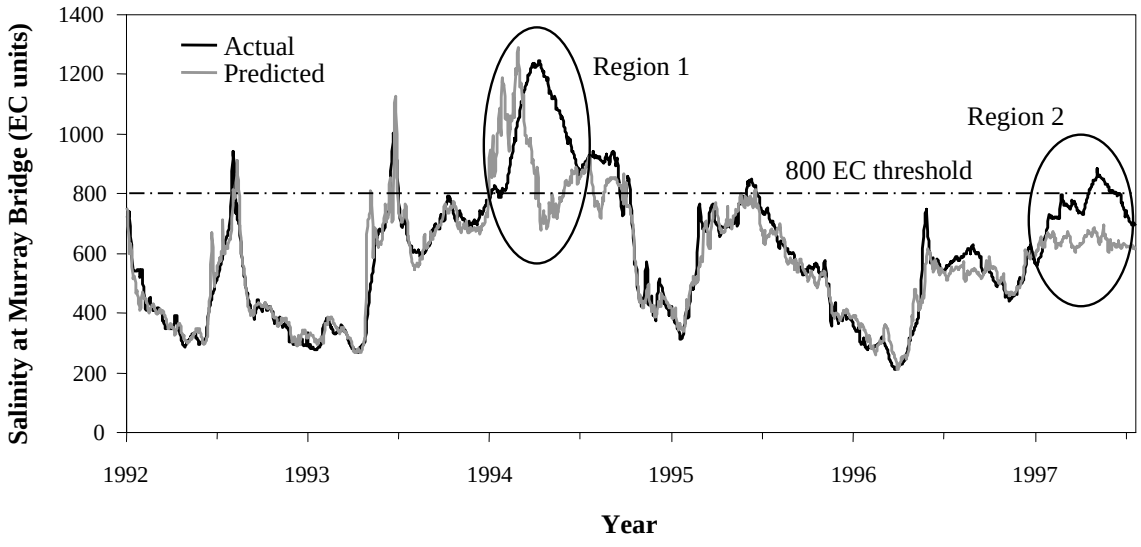

Figure 5.22 Time series plot of 14-day salinity forecasts against the "real-time forecasting" data.

5.5 BAYESIAN ANN DEVELOPMENT

The Bayesian framework for ANN development, proposed in Chapter 4, was next applied to the salinity forecasting case study, in order to compare the results obtained with those obtained in the deterministic part of the study and assess the advantages of a probabilistic modelling approach.

5.5.1 Methods

5.5.1.1 Model Selection

As the Bayesian training framework makes use of the network weights estimated using deterministic training methods, it is sensible to use the information gained from the deterministic models to guide Bayesian model selection. Since the optimal model structure was found to contain 4 hidden nodes in the deterministic part of this study, networks with $2, 4, \ldots, 10$ hidden nodes were initially considered for analysis with Bayesian techniques. The -1/2BIC, G-D and C-J evidence estimators, discussed in Section 4.3.4, were used to evaluate the evidence values of each network, in order to determine which of the initial models had the greatest posterior probability. According to the findings of the assessment carried out in Section 4.4.3.1, the -1/2BIC estimator was given the most weight in terms of model selection, while the G-D and C-J estimators were used for more comparative purposes. The marginal distributions for the hidden-output weights of the ANN model with the highest posterior probability were then inspected to determine whether any of the hidden nodes could be pruned from the model as described in Section 4.3.4.2. The same distributions were also inspected for the model containing two more hidden nodes than the ANN with the highest posterior probability, in order to verify that one or both of the two additional hidden nodes was unnecessary and could therefore be pruned from the network.

Similar to the deterministic model selection procedure carried out in this case study, the evidence values for networks either side (i.e. containing one more and one less hidden node) of the model with the highest posterior probability were also assessed. This is not strictly necessary for the Bayesian model selection framework, since it should be possible to identify how many hidden nodes are necessary by inspecting the marginal posterior distributions as described above. However, this was done in this study to check the results obtained and to further validate the proposed Bayesian model selection approach.

5.5.1.2 Training

The models were trained with the MCMC training method developed and described in Chapter 4, using a hierarchical prior distribution of the form given by (2.15). Four parallel chains were simulated, each initialised with the weights obtained for the best models developed in the deterministic part of this study. In Section 4.4, it was recommended that the hyperparameters $\sigma^2_{\mathbf{y}}$ and $\sigma^2_{\mathbf{w}}$ be fixed for an initial period $t_{\sigma^2_0} = 500$, such that the magnitude of the initial log likelihood is reduced by approximately 10%. For the $2, \ldots, 10$ hidden node ANN models developed using deterministic methods, the residual variance values $\hat{\sigma}^2_{\mathbf{y}}$ calculated at maximum log likelihood (based on scaled data) were between approximately 0.01 (10 hidden node ANN) and 0.03 (2 hidden node ANN), with corresponding log likelihood values between approximately 390 (2 hidden node ANN) and 1130 (10 hidden node ANN). Therefore $\sigma^2_{\mathbf{y}}$ was initially fixed equal to 0.04 for each MCMC simulation, which resulted in reductions in the log likelihood values between approximately 8% and 35%. For each of the models developed, the MCMC algorithm was initially run for 600,000 iterations and traces of the mean $\log p^*(\mathbf{w}|\mathbf{y})$, the mean $\log L(\mathbf{w})$ and the mean $\log p(\mathbf{w})$ densities, calculated by taking the average of the four parallel chains, were inspected to determine whether or not convergence had been achieved within this time and to determine the appropriate number of burn-in iterations t_b. Traces of the $\log p^*(\mathbf{w}|\mathbf{y})$, $\log L(\mathbf{w})$ and $\log p(\mathbf{w})$ values obtained from the individual chains were also inspected to assess convergence. A random sample of 10,000 weight vectors, simulated after convergence was achieved, were then used to calculate predictive distributions for the given input data. From these, the mean predictions and 95% prediction limits were evaluated.

5.5.1.3 Validation

The optimal model structure, as selected using the BMS approach, was validated by subjecting it to the independent testing and validation subsets, not used for training. As the test data set was not required for cross-validation under the Bayesian framework, it could have been combined with the training data to form a larger training data set. However, in order to perform a fair comparison of the Bayesian forecasts obtained with the deterministic forecasts, this was not done in this study.

The RI distributions for each input were determined by applying the modified Connection Weight Approach to the weight vectors sampled from the posterior distribution. These were then compared to the PMI-based RI estimates to evaluate whether the underlying physics had been captured by the model and to assess how uncertain the input

contributions were.

5.5.2 Results

It was found that 600,000 iterations were insufficient to achieve convergence of the MCMC simulations for all of the network sizes. Therefore, the algorithm was run for an additional 600,000 iterations for each ANN. It can be seen in Figures 5.23 and 5.24, which show traces of the mean $\log p^*(\mathbf{w}|\mathbf{y})$, $\log L(\mathbf{w})$ and $\log p(\mathbf{w})$ densities and the individual $\log p^*(\mathbf{w}|\mathbf{y})$ densities, respectively, that apart from the 6 hidden node ANN, the MCMC algorithm converged after approximately 600,000 iterations. For the 6 hidden node ANN, convergence was achieved after approximately 900,000 iterations. Therefore, the number of burn-in iterations that were discarded for the 2, 4, 8 and 10 hidden node ANNs was 600,000, while this number was 900,000 for the 6 hidden node model. The MCMC algorithm was run for an additional 300,000 iterations for the 6 hidden node ANN so that the number of weight vectors sampled after burn-in was the same for each of the models. Training took between 1.5 and 11.4 hours for the different network sizes, which was comparable to the training times taken by the SCE-UA algorithm (see Table 5.4).

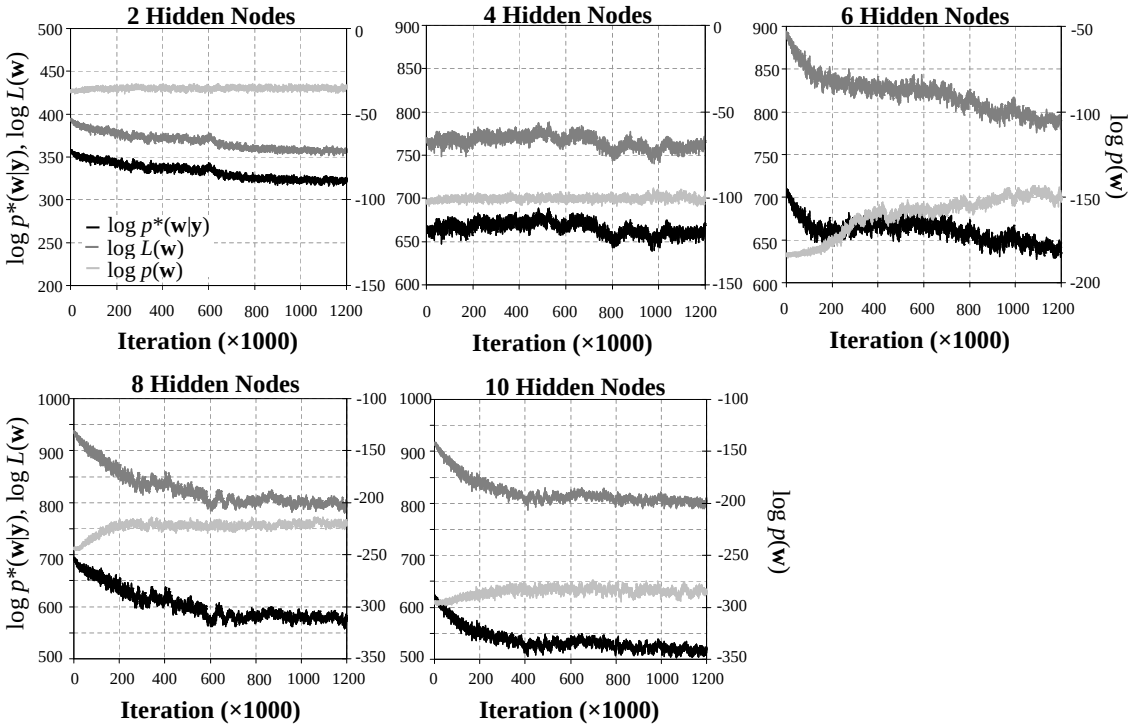

Figure 5.23 Mean $\log p^*(\mathbf{w}|\mathbf{y})$, $\log L(\mathbf{w})$ and $\log p(\mathbf{w})$ traces for the $2, 4, \ldots, 10$ hidden node ANNs.

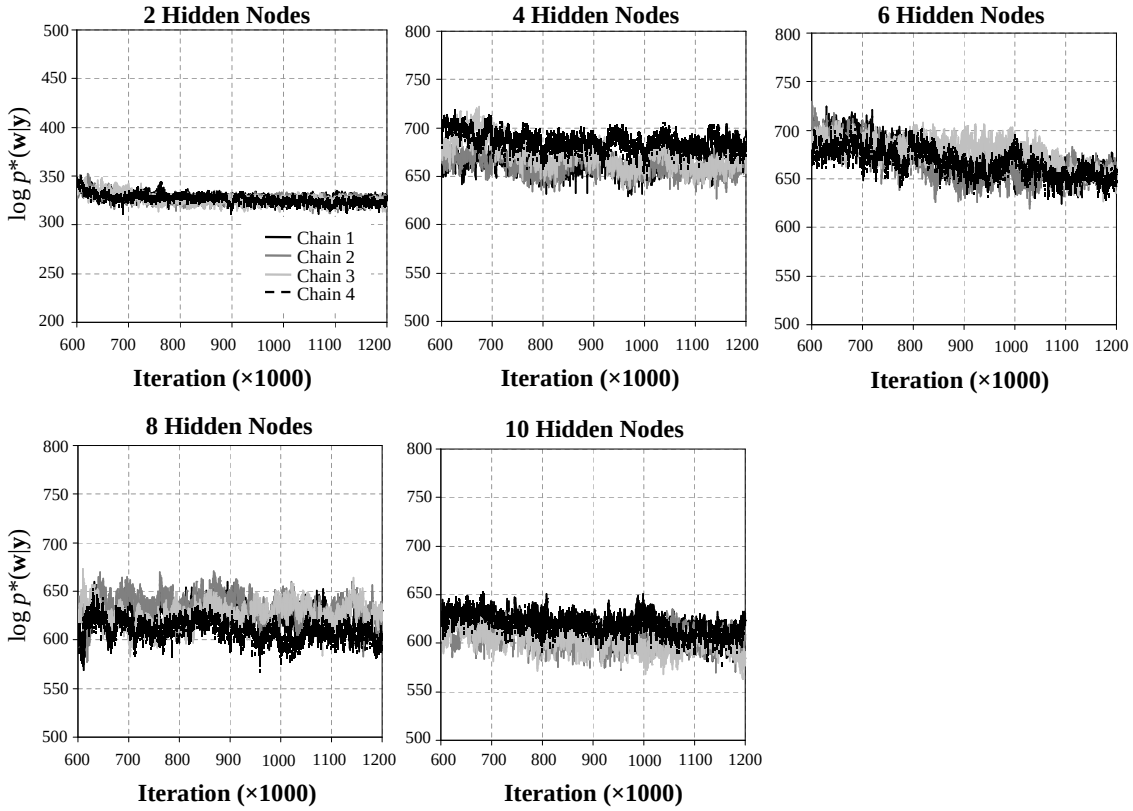

Figure 5.24 Log $p^*(\mathbf{w}|\mathbf{y})$ traces obtained from the 4 parallel MCMC chains for the $2, 4, \ldots, 10$ hidden node ANNs.

The evidence values estimated with the -1/2BIC, G-D and C-J estimators are plotted in Figure 5.25, where it can be seen that the 4 hidden node ANN had the greatest evidence according to all 3 of the estimators. The $BF_{Rank1,i}$ results, calculated based on the $-1/2\mathrm{BIC}$ evidence values and presented in Table 5.6, indicate that there is very strong evidence in favour of the 4 hidden node ANN over the other network sizes. To check this result, the marginal posterior distributions for the hidden-output weights of the 4 and 6 hidden node ANNs were inspected. These are shown in Figures 5.26 and 5.27, respectively. It can be seen that for the 4 hidden node model, all of the hidden nodes are necessary, as zero is not included within the 95% highest density regions of any of the weight distributions. It can also be seen that, for the 6 hidden node model, zero is included in the 95% highest density regions of the distributions for hidden-output weights 2 and 6 (Figure 5.27 (b) and (f)), indicating that at least one of the corresponding nodes could be pruned from the network. By inspection of the scatter plot for hidden-output weight 2 versus hidden-output weight 6, shown in Figure 5.28, it was evident that both

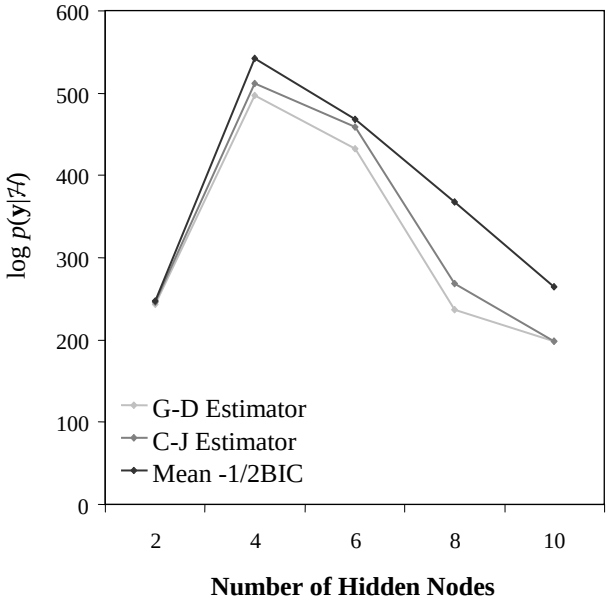

Figure 5.25 Evidence estimates for the $2, 4, \ldots, 10$ hidden node ANNs.

nodes could be pruned, as the joint distribution was found to pass through the origin; thus leaving a network with 4 hidden nodes.

To verify that the 4 hidden node ANN was indeed the optimal structure for providing 14-day salinity forecasts, networks with 3 and 5 hidden nodes were trained and assessed using the Bayesian approach. The resulting evidence values estimated with the -1/2BIC, G-D and C-J estimators are plotted in Figure 5.29, in comparison to the values of the five original models. It can again be seen that the 4 hidden node ANN had the greatest overall evidence according to all of the estimators. The marginal posterior distributions for the hidden-output weights of the 3 and 5 hidden node ANNs are shown in Figures 5.30 and 5.31, respectively. It can be seen that all of the hidden nodes are necessary for the 3 hidden node model, as none of the distributions are close to zero. However, the distribution

Table 5.6 Log Bayes Factors in favour of the highest ranked model.

Rank	Hidden Nodes	$\log_e BF$ in Favour of Rank 1 Model
1	4	–
2	6	74.690
3	8	175.286
4	10	277.341
5	2	295.372

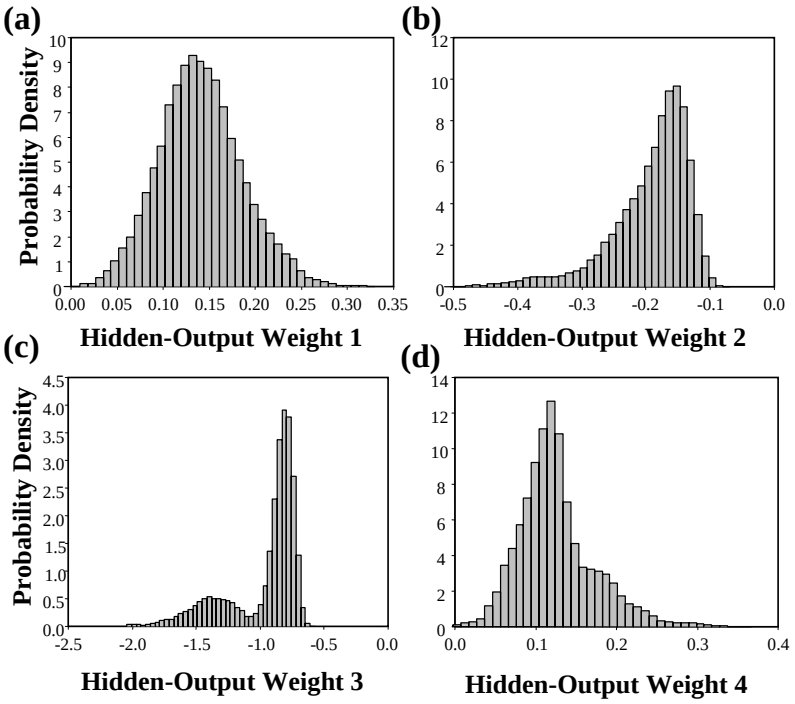

Figure 5.26 Marginal posterior hidden-output weight distributions of the 4 hidden node ANN.

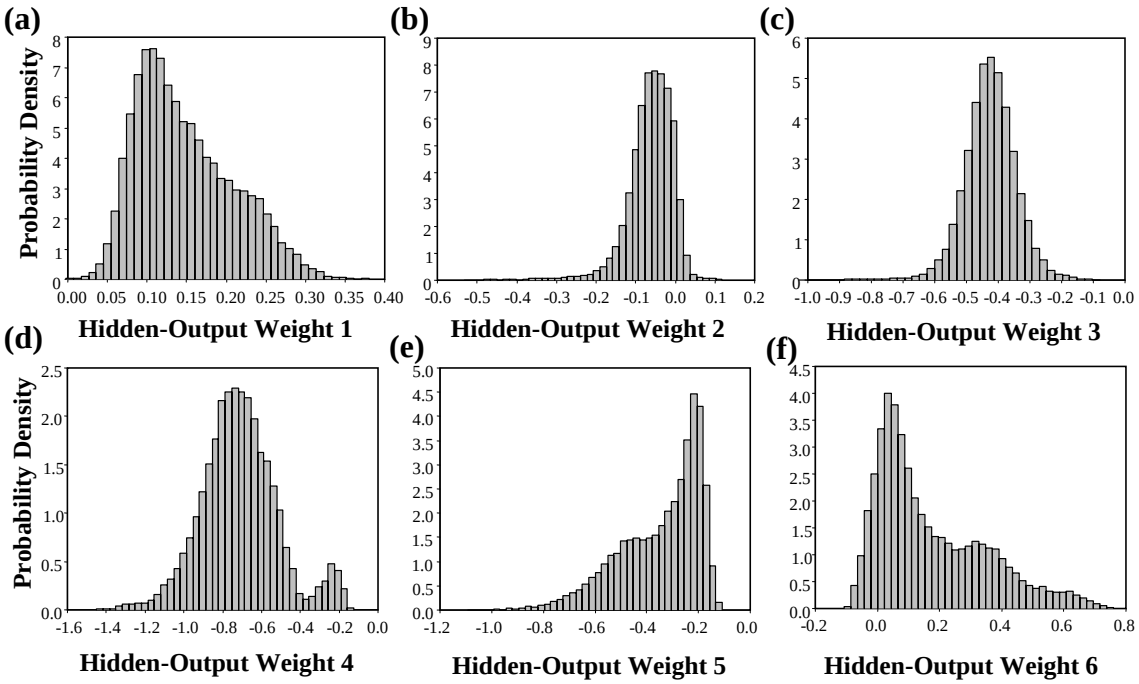

Figure 5.27 Marginal posterior hidden-output weight distributions of the 6 hidden node ANN.

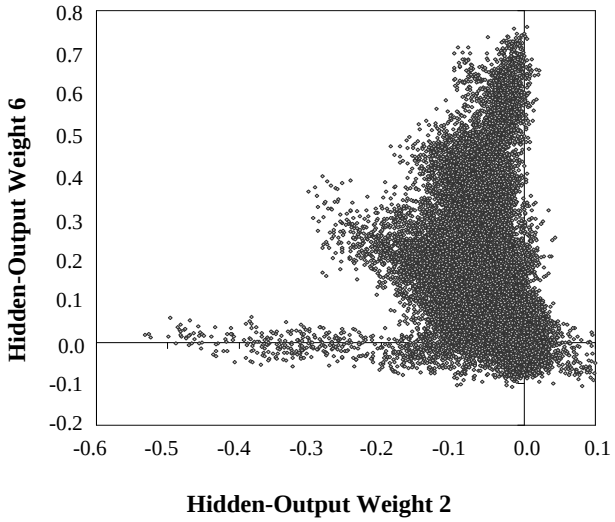

Figure 5.28 Scatter plot of hidden-output weight 2 versus hidden-output weight 6 for the 6 hidden node ANN.

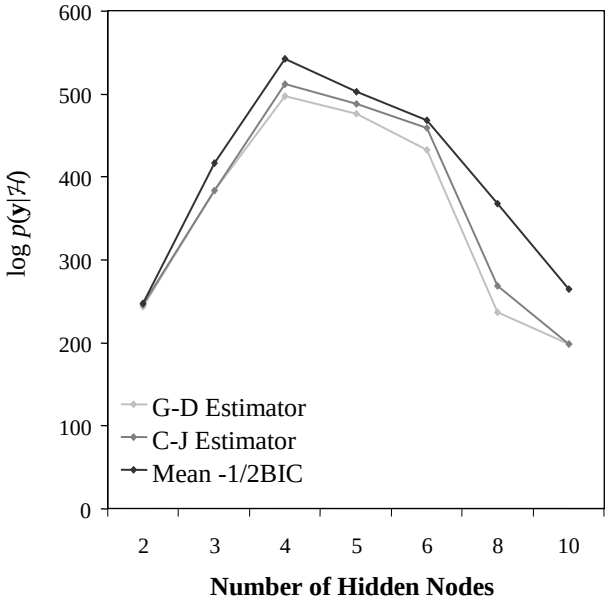

Figure 5.29 Estimated evidence values including those estimated for the 3 and 5 hidden node ANNs.

for hidden-output weight 3 (Figure 5.31 (c)) does include zero with the 95% highest density region, indicating that the corresponding hidden node is unnecessary, and hence a 4 hidden node ANN is more optimal.

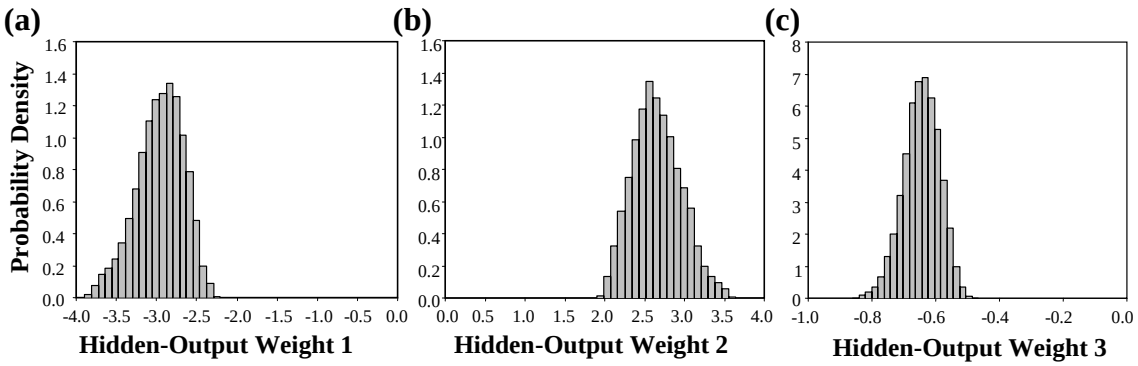

Figure 5.30 Marginal posterior hidden-output weight distributions for the 3 hidden node ANN.

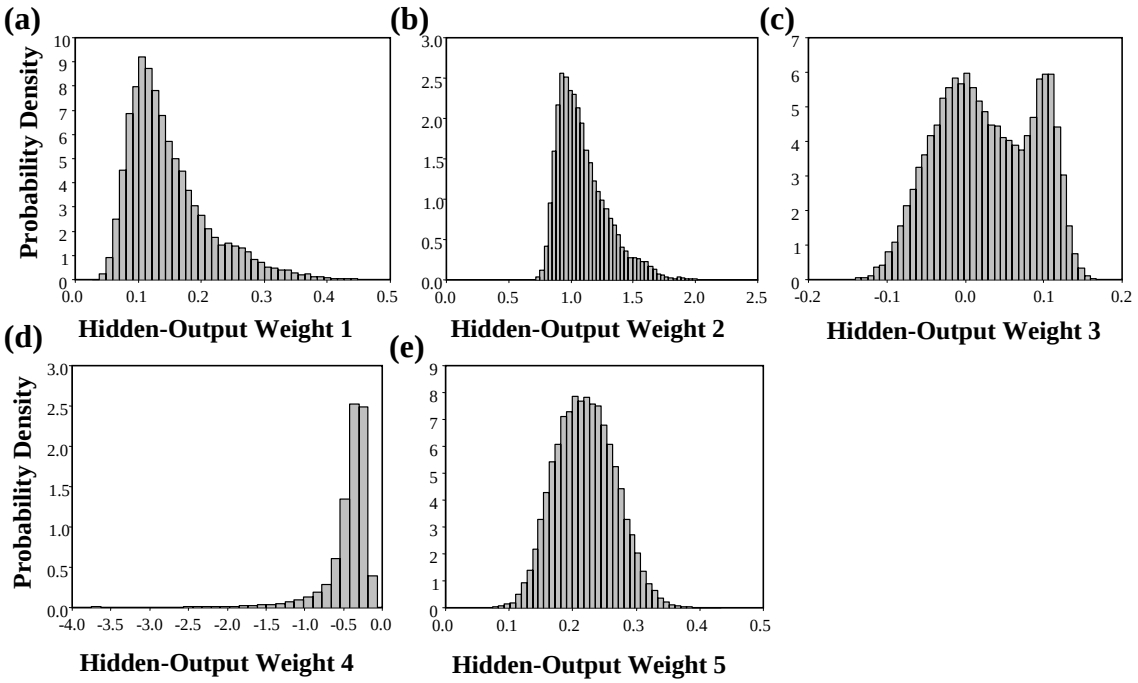

Figure 5.31 Marginal posterior hidden-output weight distributions for the 5 hidden node ANN.

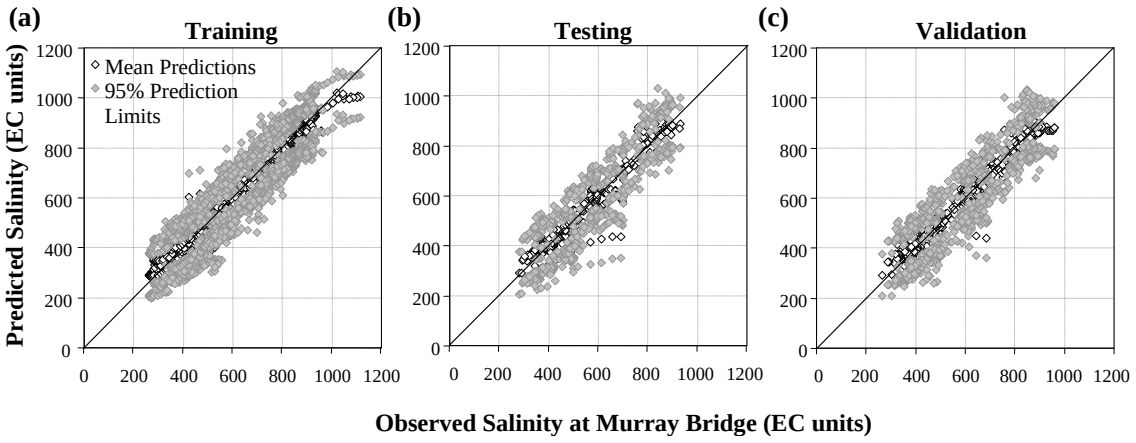

Figure 5.32 Scatter plots of the 4 hidden node ANN model mean predictions and 95% prediction limits versus the (a) training, (b) testing and (c) validation data.

Shown in Figure 5.32 are scatter plots of the mean model predictions and 95% prediction limits obtained using the 4 hidden node ANN versus the training, testing and validation data. It can be seen in this figure that the mean predictions provide a good fit to the measured data and are, in fact, very similar to the deterministic forecasts shown in Figure 5.20. Greater than 95% of the observed salinity concentrations were accounted for by the 95% prediction limits for the training, testing and validation data sets. Shown in

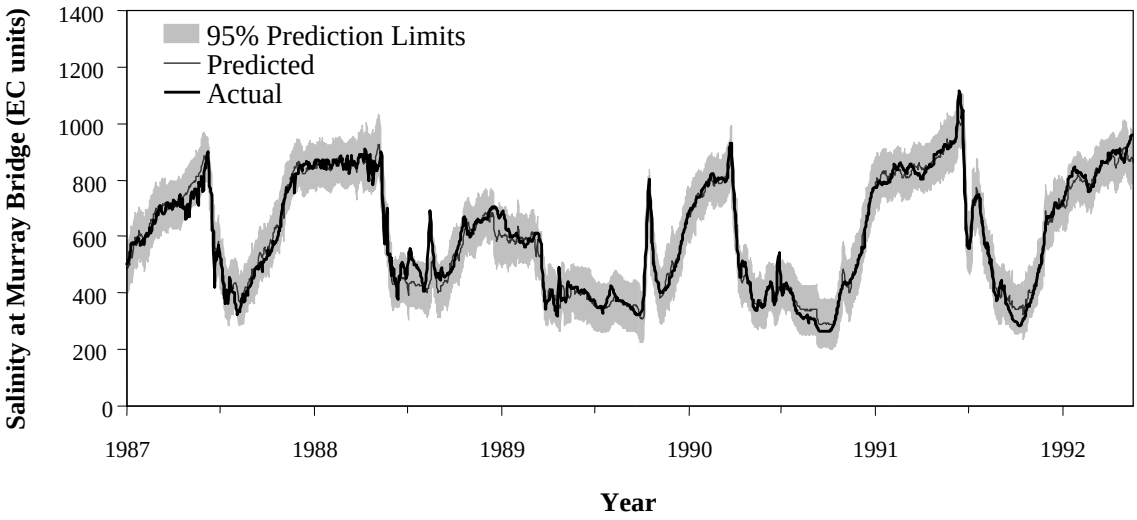

Figure 5.33 Time series plot of the mean 14-day salinity forecasts and 95% prediction limits obtained with the 4 hidden node ANN model against the "model development" data.

Figure 5.33 is a time series plot of the mean 14-day salinity forecasts and 95% prediction limits against the recombined model development data (training, testing and validation). From the width of the prediction limits, there appears to be relatively little uncertainty associated with these forecasts.

The RI distributions for each input are shown in Figure 5.34, while the minimum, maximum and mean values of these distributions are presented in Table 5.7, together with the PMI- and MI-based RI estimates and the order of input importance according to the mean model-based RI values. The mean values in Table 5.7 are relatively similar to those that were calculated based on deterministic weights in Table 5.5, and are of similar order of importance as the PMI-based RI values, and approximately the same magnitude as the MI-based estimates. While it is apparent in Figure 5.34 that most of the RI values are fairly uncertain, it is still obvious from these distributions which are the important inputs for forecasting salinity concentrations at Murray Bridge, as these inputs have RI values that are not distributed about zero. As discussed previously, all deterministic RI estimates are subject to uncertainty, including the PMI-based estimates. The RI distributions have the advantage of identifying which inputs are important predictors and those which are not, without trying to pin-point exact values for their importance to the overall model.

Table 5.7

	Rank	Model-based				
Input	Importance	Minimum	Maximum	Mean	PMI-based	MI-based
MAS_{t-1}	4	0.02	31.24	11.62	70.32	17.99
MOS_{t-60}	12	0.00	10.51	3.15	1.20	6.36
WAS_{t-1}	1	9.04	31.21	17.14	14.54	16.33
WAS_{t-43}	10	0.00	23.53	4.66	1.51	9.64
LOS_{t-25}	9	0.00	20.12	7.17	1.89	11.30
$\mathrm{L7F}_{t-1}$	3	0.04	25.20	14.78	2.44	8.73
MBL_{t-1}	11	0.00	20.21	3.76	1.41	5.21
MBL_{t-11}	10	0.00	21.13	4.44	0.93	4.78
MBL_{t-21}	8	0.00	22.11	4.69	0.76	4.50
MBL_{t-34}	13	0.00	18.06	2.79	0.79	4.13
MBL_{t-57}	7	0.00	21.57	5.00	0.83	3.69
MAL_{t-57}	6	0.00	19.14	5.03	1.33	4.02
$\mathrm{L1UL}_{t-1}$	2	0.12	31.75	15.77	2.05	3.31

Finally, the 4 hidden node ANN model was applied to the "real-time forecasting" data and a plot of the mean output time series and 95% prediction limits can be seen in Figure 5.35. For most of the forecasting period, the 95% prediction limits are approximately the same width as those calculated for the "model development" data; however, the width

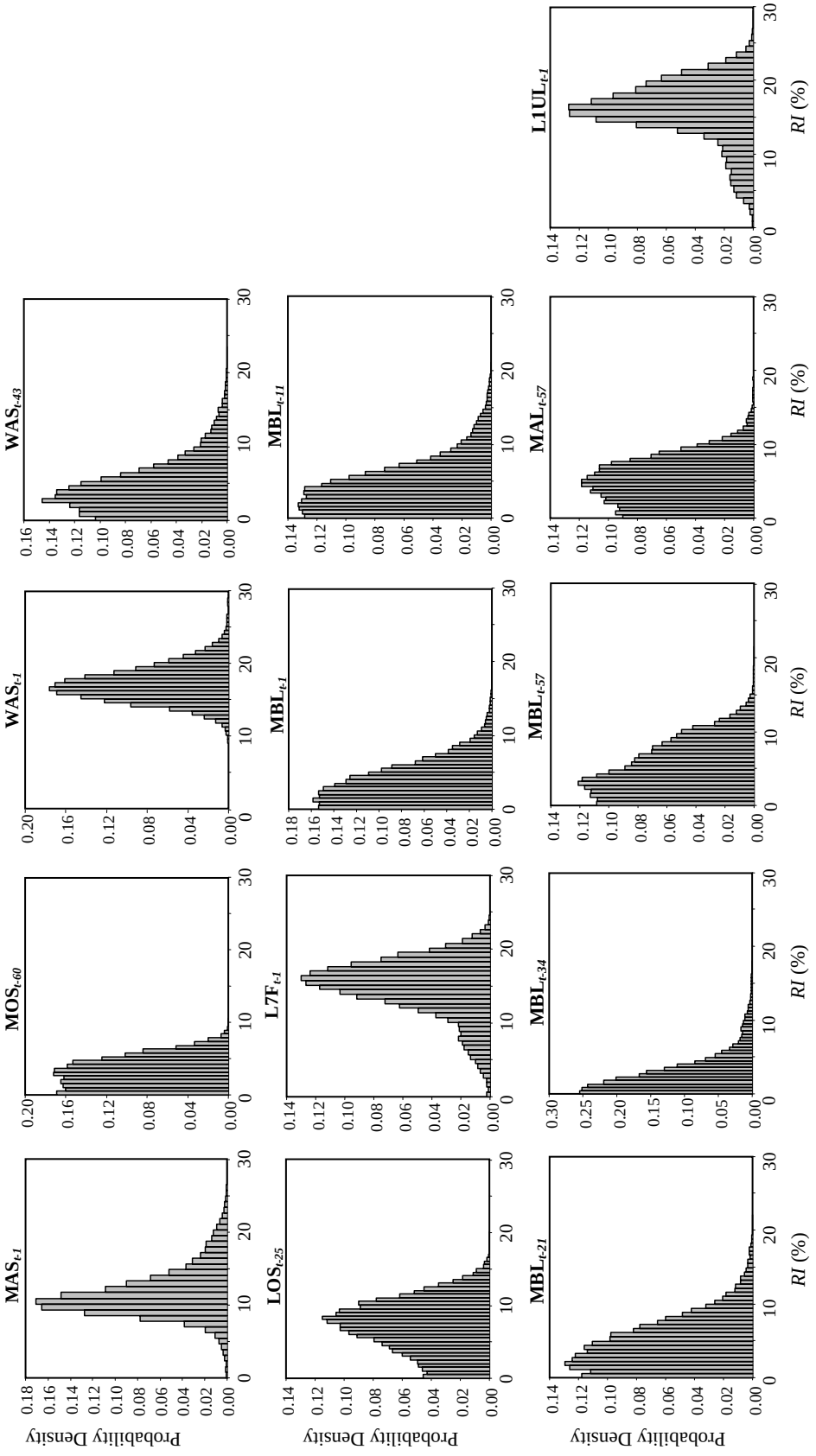

Figure 5.34 *RI* distributions for salinity inputs.

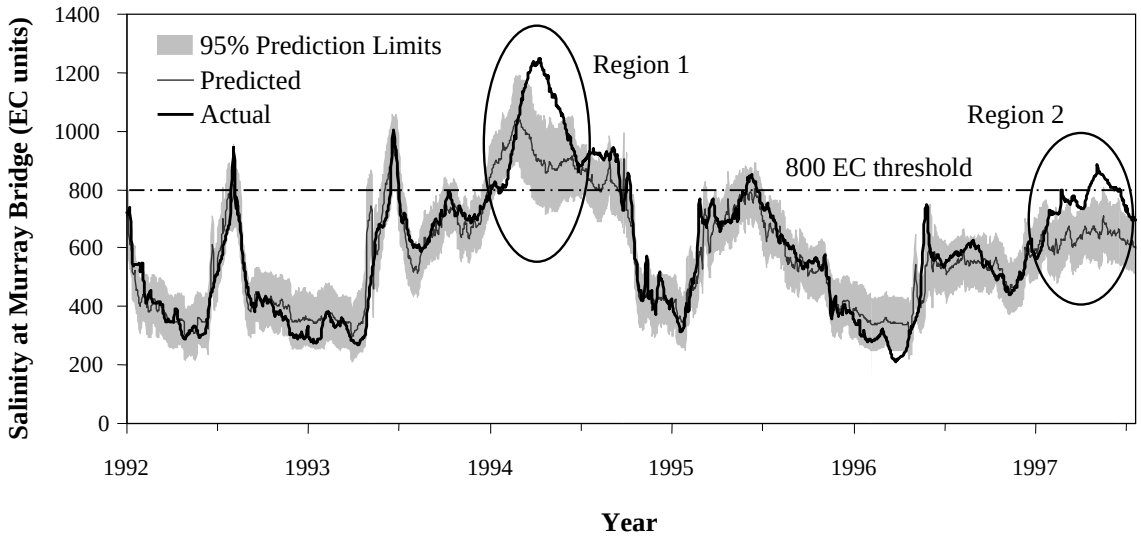

Figure 5.35 Time series plot of the mean 14-day salinity forecasts and 95% prediction limits obtained with the 4 hidden node ANN model against the "real-time forecasting" data.

of these limits increases to some extent in regions 1 and 2. During these periods the ANN has to extrapolate beyond the range of the training data and the resulting uncertainty in the forecasts due to the uncharacteristic data is reflected in the expanded prediction limits, indicating to the modeller that single valued forecasts (e.g. mean predictions) should be used with caution. Salinity levels were under-predicted in each of these regions; however, unlike the deterministic model, the mean predictions in region 1 were above the 800 EC threshold; thus providing better single valued forecasts of the actual salinity concentrations. However, it can also be seen in this figure that the 95% prediction limits failed to include all of the observed salinity data in regions 1 and 2. While this may, in part, be due to inappropriate convergence to the true posterior, uncertainty in the ANN weights is only one source of prediction uncertainty and the fact that all of the data in these regions were not accounted for by considering this source may suggest inadequacies in the model used to forecast the salinity data, possibly due to the omission of important inputs, errors in the data or limitations in the model structure. Nevertheless, the Bayesian ANN provides a significant improvement over the deterministic ANN, and apart from the two periods of uncharacteristic data, almost all data points fall within the 95% limits. If this model was used in a management situation, the manager would be able to decide, based on the information provided in the mean forecasts and in the predictions limits, what is the likelihood that the salinity concentration in the river will exceed 800 EC units, which should ultimately lead to better management of the resource.

Table 5.8 presents the results obtained using the deterministic and Bayesian ANNs in comparison to the results obtained by *Bowden et al.* (2005b). These results show that the performance of the Bayesian model is significantly better than the other models in the real-time forecasting scenario. This highlights the importance of accounting for the entire range of plausible weight vectors when making predictions, rather than relying on the single weight vector that provides the best fit to the training data. By estimating the posterior weight distribution, the Bayesian ANN has achieved a more generalised mapping of the underlying relationship, which is influenced less by the minimum error of the training data and influenced more broadly by the overall information contained in the data. This can be seen by the fact that the Bayesian ANN had the poorest performance on the model development data; yet, as seen in the real-time forecasting results, enabled better extrapolation. It can also be seen in this table that both the deterministic and Bayesian ANNs developed in this research had similar performance to the 32 hidden node ANN developed by *Bowden et al.* (2005b), even though these models contained 420 fewer weights. This highlights the importance of appropriate model selection techniques, as smaller models are less susceptible to overfitting, more efficient to train, have a smaller degree of uncertainty associated with the weights and resulting predictions and are easier to interpret.

Table 5.8 Performance of 4 hidden node ANN salinity forecasting model developed using deterministic and Bayesian methods.

Performance	Model Development Data			Real-Time
Measure	Train	Test	Validation	Forecasting Data
Deterministic				
MAE	19.3	26.1	29.5	63.0
RMSE	26.7	39.0	39.3	105.7
CE	0.98	0.95	0.96	0.79
Bayesian				
MAE	23.3	29.5	30.6	62.1
RMSE	31.9	43.6	41.2	88.4
CE	0.98	0.94	0.95	0.85
Bowden et al. (2005b)				
RMSE	29.3	30.8	34.0	95.0

5.6 CONCLUSIONS

The results of the salinity forecasting case study presented in this chapter highlight the importance of accounting for the uncertainty associated with ANN predictions and demonstrate the advantages of the proposed Bayesian training approach over deterministic ANN development techniques. While the performance of the ANN model developed using the Bayesian approach was similar to that of a deterministic ANN in an interpolative context, it was shown that the Bayesian ANN was more robust in a real-time forecasting scenario, particularly when the model was required to extrapolate. Not only were the average forecasts obtained using the Bayesian ANN an improvement over the single valued forecasts obtained using the deterministic ANN, but prediction limits, indicating the quality of the forecasts, were produced using the Bayesian approach, which was shown to be particularly important in situations when forecasts were made outside the range of the calibration data.

The deterministic and Bayesian model selection methods were both able to select the most appropriate structure; however, it was shown that prevention of overfitting using a test data set was still necessary for the deterministic approach. On the other hand, a test data set is not required for the Bayesian approach, which means that a greater amount of data, and hence information, is available for training the model.

It was shown that even with modifications to the SCE-UA parameters, the resulting training times can still be excessively long in complex real-world cases, involving numerous inputs. In this research, the aim of using this algorithm for training was to provide the best deterministic results for comparison with the Bayesian approach, in order to properly evaluate the advantages of using the proposed Bayesian methodology. However, if the Bayesian framework was to be used for developing an ANN model in a high dimensional real-world application, it is recommended that a simpler, gradient-based training algorithm, such as backpropagation, be used to obtain rough initial estimates of the mode of the posterior distribution.