Predicting transit times and flow concentrations of contaminants through a metropolitan sewage system

J. Barry　　　　J. Crawford　　　　P. Airey [*]

Abstract

Neural Networks were used to build a model that would predict the transit times and the levels of trace amounts of radioactive components at the outlet of a sewage treatment plant. The Neural Networks are

---

[*]Australian Nuclear Science and Technology Organisation Private Mail Bag 1, Menai 2234, Australia. mailto:jmb@ansto.gov.au, mailto:jc@ansto.gov.au, mailto:pla@ansto.gov.au

compared with more conventional approximation methods as a tool for prediction.

Contents

1   Introduction

The third author challenged the others to examine the potential of using Neural Networks to model complex dynamic behaviour in environmental systems. The motivation was to experiment with a *black box* technique instead of other physical modelling systems which were considered too costly to develop and implement. The following problem was considered sufficiently complex to be of use in evaluating the value of Neural Networks in modelling such systems.

ANSTO routinely discharges low levels of radionuclides, including tritium to the Cronulla sewage system under the appropriate authorisation. Under its Health, Safety and Environment Policy, the Organisation is committed to undertaking its activities consistent with national and international best practice. This involves a commitment to continuous improvement.

To effect continuous improvement, it is important for ANSTO to be in a position to predict the transit time of the discharge before it reaches the treatment plant and the levels of radioactivity as it passes through the plant under a wide range of conditions.

2   Physical system

The sewage system to which ANSTO is linked is difficult to model as it consists of many conduits, holding locations and intermediate pumping stations

of differing rates and delay times. In addition to normal household and industrial discharges there are many other environmental and operational factors that are difficult to account for in a conventional modelling approach. The complexities of the system were not incorporated directly in the models.

Tritium (as HTO) was modelled in this study as it behaves as water and is not retarded by adsorption on the pipeline walls or by the sewage sludge. The discharges from the ANSTO holding tanks are made for periods of between 3 to 7 hours. The tritium concentration is known and constant due to a mixing process in the tanks prior to discharge. The effluent is discharged into the sewage system at a reasonably constant rate. In the models it is represented by hourly varying pulses. The ANSTO contribution to the system is approximately 4% of the water volume carried by the sewage system. Samples at the Cronulla Sewage Treatment Plant (CSTP) are taken at hourly intervals, starting before the effluent reaches the plant and the measurements are continued until the material has passed through. The samples are returned to the laboratory for tritium assay. Because of possible losses from the system due to damaged pipes and possible retention of portion of the release in intermediate stations, the system is not necessarily conservative. Consequently no attempt was made to force conservation in the models.

Two empirical modelling approaches were adopted. A neural network was developed in a typical *black box* approach to predict the flow times and concentrations so as to avoid the need for finding a suitable mathematical approximation function. The network was trained with data recorded from a limited number of releases from ANSTO. In a more traditional response

model, an approximation function was developed and parametrised through minimisation of least squares error component. As in the case of the neural network, specific details of the sewage network was not included in the model.

3   Conventional model

With the conventional modelling approach, it is advisable to consider possible mathematical mechanisms that may describe the physical system. If the time varying tritium load from the ANSTO facility is represented by $I(t)$ and the tritium load measured at the CSTP by $C_{out}(t)$, then an integral equation of the form

$$C_{out}(t) = \int_0^t I(s)f(t,s))ds \qquad (1)$$

may be used to relate the two quantities provided the kernel function $f(t,s)$ can be determined. We persist with a kernel function of this form, however, the situation may well be more complex for $f(s,t,I(s))$ may turn out to be more appropriate form. Equation 1 is a Fredholm Integral equation of the first kind, where normally the $C_{out}(t)$ and $f(s,t)$ are known and $I(t)$ is sought. These problems can be difficult to solve because they are typically ill-posed [2]. The ill-posed nature of the problem shows up in the recovery of the kernel functions because the kernel does not always depend continuously on the data.

If one fails to recognise the integral equation formulation either in setting up a neural network solution or in a standard least squares approximation, then there can be important implications for the solutions obtained.

One may set up an approximation to the physical system or attempt to discretise the integral equation through

$$C_{out}(t) = \sum_{i=0}^{n} I(t_i)g_i(t - t_{delay}, \mathbf{a}^i), \tag{2}$$

where the inputs are treated as $n + 1$ pulses at times t_i generally on an hourly basis, when the discharge flow rates were measured and recorded. The output loads are related to the input loads through the convolution function $g_i(t - t_{delay}, \mathbf{a}^i)$, where t_{delay} is the time of flight for a particular stage of the release. In the present model, t_{delay} is assumed constant for every stage of a particular release. The functional form is parametrised by sets of parameters $\mathbf{a}^i$.

As it stands, there are a set of $n + 1$ convolution functions to be determined for each release used to determine the model. This would require the computation of a set of parameters $\mathbf{a}^i$ for each release. As part of the simplification it is assumed that the same parameters apply for every release, this corresponds with the decision not to use the $f(s, t, I(s))$ form of the kernel.

3.1   Mathematical basis function

The convolution function $g(t, \mathbf{a}^i)$ of equation (2), is defined with linear hat basis functions v_i

$$g(t - t_{delay}) = \sum_{j=1}^{r} a_j v_j(t - t_{delay}), \tag{3}$$

based on r nodal points in time $T_1 < T_2 < \ldots < T_r$. The points are chosen at r suitable time displacements from $t - t_{delay}$. Using $t' = t - t_{delay}$, the basis functions v_i are defined:

$$v_1 = \frac{t' - t_0}{T_1 - t_0} \qquad \text{for} \qquad t' < T_1, \tag{4}$$

$$v_j = \frac{t' - T_{j-1}}{T_j - T_{j-1}} \qquad \text{for} \qquad T_{j-1} \leq t' \leq T_j \quad j = 2, 3, \ldots, r, \tag{5}$$

$$v_j = \frac{T_{j+1} - t'}{T_{j+1} - T_j} \qquad \text{for} \qquad T_j \leq t' \leq T_{j+1}, \quad j = 1, 2, \ldots, r-1, \tag{6}$$

$$= 0 \qquad \text{elsewhere;} \tag{7}$$

$$v_r = 1 - \frac{t' - T_r}{t_\infty - T_r}, \quad \text{for} \qquad t' > T_r, \tag{8}$$

where t_0 is the time at which the release of effluent from ANSTO commences and t_∞ is the time after the effect of any discharge passes through the system.

In parametrising the model all releases are considered simultaneously and the parameters are recovered by minimising

$$\sum_{k=1}^{p} \sum_{j=1}^{m_k} \left[\frac{C_{out-observed}^k(t_j^k) - \sum_{i=1}^{n_k} \sum_{l=1}^{r} I^k(t_i^k) \mathbf{a}_l \upsilon_l(t_j^k - t_{delay}^k)}{\sigma_{j,k}} \right]^2, \quad (9)$$

where p different releases are used; for the kth release there are m_k observations at the outfall, n_k input observations and a separate delay constant, t_{delay}^k for each of the p releases. Variance estimate $\sigma_{j,k}$ were not available, so unit variance was assumed. Minimisation was easily obtained with a quasi-Newton algorithm. When the model is used in a predictive mode, t_{delay}^k is assumed to be linearly related to the time when the release originated.

3.2   Stability

The model was first parametrised using $r = 7$ nodal points. The model was able to reproduce the observed loads at the CSTP and the delay times rather well. However, a sensitivity test on the parameters was undertaken and as anticipated, large fluctuations in the recovered parameters were detected for small perturbations of the underpinning data; as typically experienced in ill-posed problems of this form. In a move to reduce these fluctuations, two approaches were adopted. Tikhonov regularisation[2] was implemented and alternatively, the number of nodal points allowed in the basis functions were restricted. The second proved effective and for its inherent simplicity was

the method ultimately adopted. To evaluate the value of the model further, the sensitivity of each parameter to data error was tested by Monte Carlo simulations for $r = 3, 5$ and 7. A 1.0% perturbation, randomly normally distributed was applied to the input data. There were four releases used in the parametrisation, hence four values for t_{delay} were recovered as well as the spatial parameters **a**. The standard deviation of each parameter was estimated in the Monte Carlo simulation and each is expressed as a percentage of the recovered parameter in Table 1, where an 'x' in the table indicates the optimisation algorithm consistently drove that recovered parameter to zero. For $r > 3$ the standard deviations of the parameters are so large as to bring into doubt the reliability of the model, even though the model predictions match well the observed loads.

4   Neural network model

While the Artificial Neural Networks (ANNs) [1], [4] approach is similar to the conventional optimisation approach (in that an empirical model is derived based on the data acquired from past experience) the advantage is that one does not need to define the form of the function. Two approaches were considered in representing the problem using ANNs. In both approaches the outputs of the ANN are the measured hourly concentrations at the outfall from the treatment plant. Using this approach, the hourly concentrations at

TABLE 1: Standard deviation of each convolution function parameter, expressed as a percentage of the parameter.

parameter	$r = 3$	$r = 5$	$r = 7$
a_1	0.3	10.	19.
a_2	0.	272.	322.
a_3	3.2	5.7	88.
a_4		x	3.1
a_5		21.9	x
a_6			x
a_7			31.
t_{delay1}	0.00	4.2	0.03
t_{delay2}	0.08	1.5	18.
t_{delay3}	0.06	0.00	.006
t_{delay4}	0.09	6.1	6.

the CSTP, $C_{out}(t_j)$, for a given release can be represented as follows:

$$C_{out}(t_j) = f_{t_j}(I; W), \tag{10}$$

where f_{t_j} are non-linear mapping functions, W is a vector of adjustable parameters or *weights*, and I is a vector describing the release into the system.

The measured hourly discharge concentrations form the vector I in the first approach while parameters describing the discharge concentrations form I in the second approach. The first approach will not be discussed here, as it was considered restrictive, in that the duration of a discharge and the time of day when this occurred were fixed. Increasing the operational hours of discharges would result in a much larger network and hence a requirement of a larger training set of observations. In this first approach, while the observed measurements were able to be learned and reproduced fairly accurately, the predictive nature of the network was inferior. This is not surprising given the size of the network (a 8–25 network, with 200 weights to be recovered) and limited amount of training data available. The second approach is described in subsequent sections.

4.1   ANN model formulation

The ANN model is developed assuming a constant discharge rate and concentration. The inputs to the ANN (components of vector I) are the start time of the release, the duration of the release, the tritium load of the discharge

and the water discharge rate into the system. The outputs of the ANN are the concentrations at the outfall from the treatment plant.

A feed-forward ANN, using *supervised learning* [5] was considered appropriate for this problem and a package [6] was obtained for use. The package offers a number of activation functions and learning algorithms. A Gaussian activation function was chosen and the quickprop [3] learning algorithm.

In choosing a suitable ANN architecture, a number of experiments were carried out with hidden layers, however, it was found that adding a hidden layer did not improve the performance. In most cases, the predictive capability of the network was improved when the hidden layer was removed; hence a network without hidden layers was considered. To reduce the number of weights to be recovered, the measurements at the outfall were limited between 6 and 21 hours after the release, resulting in a 4–16 network.

4.2   Stability of the ANN model

To evaluate the value of the neural network model, a number of studies were carried out as described in [5], e.g. limiting the magnitude of the weights and ensuring that the distribution of weights formed a centred histogram. These conditions were satisfied, however they are not reported.

In addition, for this comparison, a sensitivity test on the weights was undertaken as described in Section 3.2, i.e. a 1.0% random normal perturba-

tion on the input was applied. The standard deviation of each weight was estimated using Monte Carlo simulation. The standard deviation of each weight, expressed as a percentage of the recovered weight, varied from 0.1% to 20.0%, with four outliers of 27%, 63%, 40% and 23%. The larger standard deviations corresponded to the sections of the curves with sharp rise and fall.

5   Evaluation of the models

To parametrise both models, data from four releases were used, these were the first four of six sets of ANSTO data recorded in chronological order. The remaining two were retained to test the predictive worth of the two models. The conventional and neural network models obtained for a typical training set are shown in Figures 1 and 2, where the observed and predicted loads are displayed. Superimposed on the predicted loads is an estimate of the standard deviation obtained through a Monte Carlo simulation, where a random normally distributed perturbation of 10.0% of the release load was applied. In each case the predictions produced are considered useful. The models give a reasonable estimate of the time of arrival of the effluent at Potter Point, the level of the load and the time it takes to pass through the plant.

The predictive power of each approach is shown for the first test release in Figures 3 and 4. In this test case there were significant differences in the discharge regime to those used in *training* the models. Consequently the models are being tested in a truly predictive mode. The more conventional
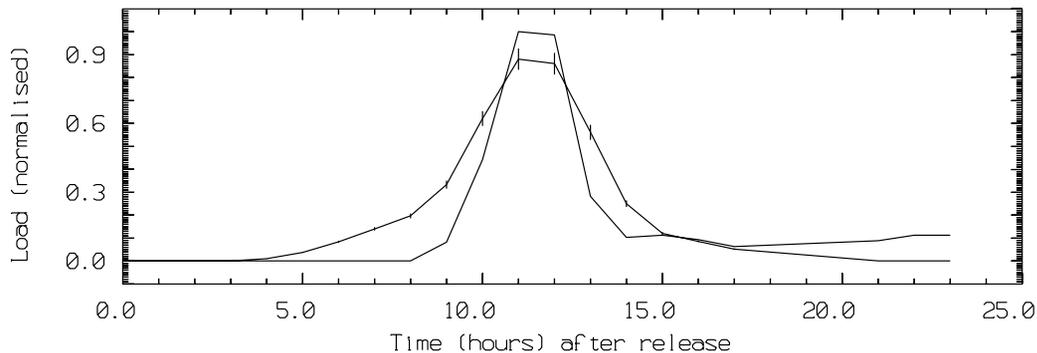
FIGURE 1: Observed and predicted loads for a release used in developing the conventional model, standard deviations are shown on prediction.
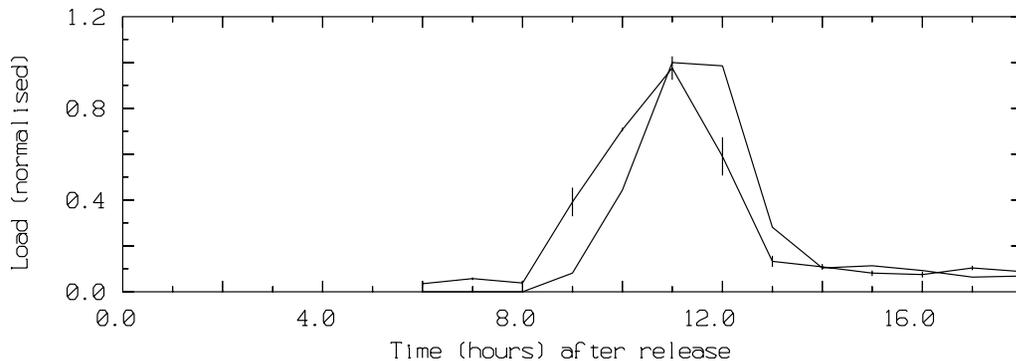
FIGURE 2: Observed and predicted loads for a release used in developing the ANN model, standard deviations are shown on prediction.
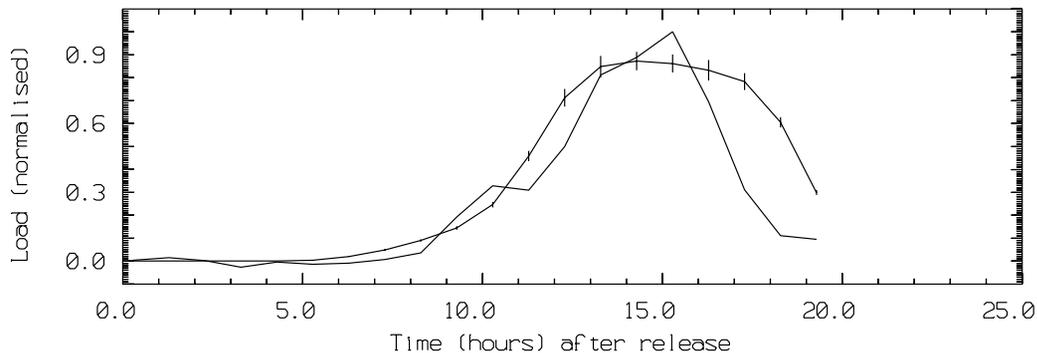
FIGURE 3: Observed and predicted loads for a release used in testing the prediction of the conventional model, standard deviations are shown on prediction.

model is a superior predictive tool in this instance, due no doubt to the regularisation forced through the reduced kernel function.

In a second test case, the volume discharged from ANSTO were considerably greater, however, the results are not shown. Unfortunately in this case, while the predicted arrival time is satisfactory, the model predicts a shorter than actual passage through the plant. This is offset with a correspondingly higher predicted load in the plant for the shorter period, which brings into question the use of empirically derived predictive models for situations that have marked differences to those for which they are parametrised. Such a result is not surprising where the training data is limited. An important issue is raised which cannot be answered until more release data is available;
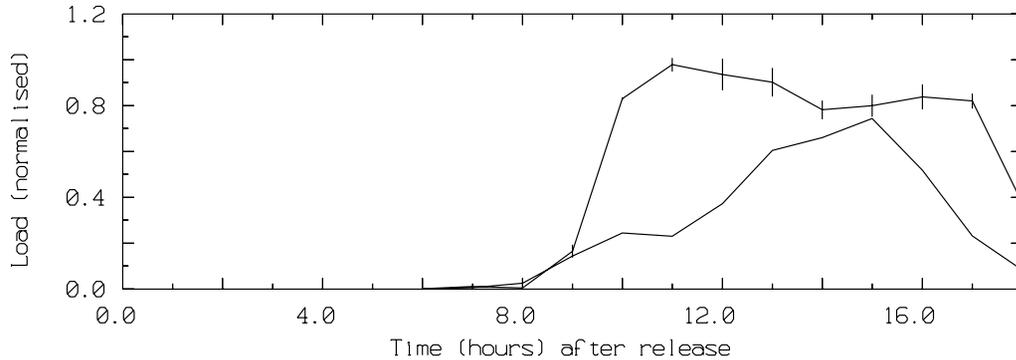
FIGURE 4: Observed and predicted loads for a release used in testing the prediction of the ANN model, standard deviations are shown on prediction.

is it possible to do better with the current empirical models with additional training data, or do the models require revision to accommodatechanged circumstances? In the sewage system there are countless extraneous factors which have not been accommodated in either approach. Their absence was a driving factor behind the use of neural networks, however, the ANN has not proven to be superior in this study.

References

[1] C.M. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford, 1997. C208

[2] H.W. Engl, M. Hanke, and A. Neubauer. *Regularisation of Inverse Problems*. Kluwer Academic, Dordrecht, Boston and London, 1996. C204, C207

[3] S.E. Fahlman. Faster-learning variations on back-propagation: an empirical study. In D. Toretzky, G.E. Hinton, and T.J. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, pages 38–51, San Mateo, CA, Morgan Kaufmann. C211

[4] W.S. Sarle. *FAQ*, available from `ftp://ftp.sas.com/pub/neural/FAQ.htm`. C208

[5] K.S. Swingler. *Applying Neural Networks*. Academic Press, London, 1996. C211, C211

[6] D.R. Tveter. *Professional Version Basis of AI Backprop*, `mailto:drt@christianliving.net`. C211