Ocean Based Water Allocation Forecasts Using an Artificial Intelligence Approach

Khan S[1], Dassanayake D[2] and Rana T[2]

[1]Charles Sturt University and CSIRO Land and Water, School of Science and Tech, Locked Bag 588, Wagga Wagga, NSW 2678, Australia, Shahbaz.khan@csiro.au

[2]CSIRO Land and Water, Wagga Wagga, NSW, Australia

Keywords: *Water allocation; sea surface temperature (SST); southern oscillation index (SOI); inflows; Blowering and Burrinjuk Dams; neuron; hidden layer; sensitivity.*

EXTENDED ABSTRACT

The initial general security water allocation announcements for water users in the Murrumbidgee Valley are made during July/August. The initial water allocation announcement is very conservative as the allocations are based on storage levels and historic minimum inflows to dams during the irrigation seasons. There is a greater than 99% chance that the more water allocations would be available as the season proceeds. Water availability during the cropping season is a major factor influencing planting decisions made by irrigators and can have a major bearing on the financial viability and irrigation efficiency of irrigation areas. Increased knowledge on the likely end-of season allocation can assist in minimising cropping risk and can help optimise farm returns and achieve better irrigation efficiencies.

In this paper a water allocation prediction framework for the Murrumbidgee Catchment is described which learns from results of detailed hydrological models and ocean based climate forecasts to predict water allocations using Artificial Neural Network (ANN) method. The ANN was successfully trained using hydrological modelling data from 1891 to 2003, to forecast water allocation for October and January months starting with a start of the season allocation in August. Subsequently Sea Surface Temperature (SST) and Southern Oscillation Index (SOI) data were added to gain greater lead times for forecasts. This network learnt well with different network parameters to that of previous networks and significantly correlated SST and SOI with water allocation. The interactive model works in a risk management context by providing probability of water allocation for the prediction month using historic data and/or with the incorporation of SST/SOI information from the previous months.

The model has been validated by forecasting January allocations from 1999 to 2005. The SST and SOI based predictions show good correspondence with the actual announced January general security allocations. Table A1 shows a comparison of historic announced and model predicted general security allocations using the ANN trained on the basis of historic data and SOI. A comparison of historic data and SOI based data show that at 50% risk factor the SOI based model results are very close to the actual announced January general security water allocations.

SST incorporated ANN model overestimated January water allocations for the 2002-2004 period. This may be due to exceptionally low starting water allocations and borrowing of water from the future years which was outside the training data sets.

Table A1. January Water Allocations Predictions Based on Historic Data and SOI

SOI Incorporated						
	Actual Water Allocation %		Model Predictions of January Allocation			
			Risk Factor			Difference
Year	August	January Next Year	25%	50%	75%	@ 50% risk
1999	50	73	50	65	84	-8.0
2000	59	90	62	90	87	-0.2
2001	47	72	47	57	82	-15.1
2002	38	38	38	45	76	7.2
2003	17	41	20	39	69	-1.7
2004	20	39	21	40	71	1.4
2005	21		21	40	71	

This study has shown that long term hydrologic simulation based water allocations at the start of the irrigation season and incorporation of a risk factor could be utilised to forecast water allocation at the end of peak irrigation demand season. Furthermore, SST, SOI and SST/SOI incorporated ANN models have significantly shown the capability to forecast end of the irrigation demand season water allocation.

1. INTRODUCTION

In a given year irrigation water allocations are made as a percentage of licensed entitlements based on storage levels in dams and actual inflows to dams. Since rainfall in a catchment can be quite variable, the initial allocation announcements at the beginning of the water year are very conservative and are based on the storage condition of dams and lowest recorded inflows (1 in 100 years) to dams. The initial water allocations do not take into account seasonal forecasts of likely inflows into dams or unregulated flows downstream of dams over the forthcoming irrigation season. Water availability is a major factor influencing cropping decisions made by irrigators and can have a major bearing on the financial viability and irrigation efficiency of irrigation areas. Increased knowledge on the likely end-of season allocation can assist in minimising cropping risk and therefore help optimise farm returns and achieve better irrigation efficiencies.

Due to complexity and the non-linearity of the allocation environments and impossibility of building linear relationship between water allocations of winter and summer periods Artificial Neural Network (ANN) method was selected to build up model applications to relate these problematic water allocations.

ANN is a proven technology that has solved non-linear problems in many applications. These networks are collections of mathematical models that emulate some of the observed properties of biological nervous systems. Therefore it learns from past data to get adapted to the system and predict for future. Its adaptive feature provides facility of accommodating new input parameters or more or less data points and getting adapted to the new situation. Therefore ANN is a promising technology to develop relationship between non-linear problematic water allocations in different seasons.

2. ARTIFICIAL NEURAL NETWORK

Artificial Neural Network (ANN), being a simple model of biological neuron system, is a collection of mathematical models that emulate some of the observed properties of biological nervous systems such as adaptive learning from historic data to seek data patterns and predict for the future. ANN are capable of storing data as patterns to recall and recognise them in a later stage like brain does.

Due to ANN's capability of handling non-linear relationships, it is suitable for complex applications such as forecasting water allocations, industrial control systems, financial forecasting pattern and voice recognition, and health sector, where linear relationships do not exist. 'Neural network practitioners generally tackle more complex problems, the dimensionality of the models tends to be much higher, and methodologies are hand tailored to particular applications' (Holger et al, 2000).

In ANN there are different network topologies and two learning modes are referred as supervised and unsupervised. In order to make use of an ANN network, it has to be trained with available inputs and outputs. When it completes its learning it can be assigned to forecast for unseen inputs to predict unavailable outputs. Non-linearity within input and output data sets is solved by introducing hidden layers into the network.

3. TRAINING NETWORKS AND DEVELOPMENT OF MODELS

3.1. Inputs and Input-output relations.

The month of August has been chosen to represent initial general security allocation month and January is selected to represent end of major water demand period. Hence one of the model inputs has been August water allocation targeting January allocation. It has been found that there is a high level of correlation between sea surface temperature and inflows (Khan et. al, 2004) to the inflows to the Blowering and Burrinjuck dams, through which water is supplied to the Murrumbidgee region. In the aforementioned study correlation between the SST and inflows to dams were calculated for each grid point of a global mesh of (2^o x 2^o) on a monthly, three monthly and seasonal basis, with lag time of up to 2 years. The Sea Surface Temperature (SST) datasets were downloaded from the National Climatic Data Center, Asheville, North Carolina. The August and January water allocation levels for the Murrumbidgee Valley with today's environmental flow rules for the years 1890–1999, were based on DLWC's IQQM model runs. This study has shown that the January sea surface temperatures of the cluster points in Table 1 (SST1) and as shown in Figure 1 have been best correlated with inflows to the Blowering dam whilst the January sea surface temperatures of the cluster points in Table 2 (SST2) have been best correlated with inflows to the Burrinjuck dam.

Table 1. Related cluster points to inflows to Blowering Dam

Blowering Dam	
(-52,72)	(-54,64)
(-52,70)	(-54,62)
(-52,68)	
(-52,66)	
(-52,64)	

Table 2. Related cluster points to inflows to Burrinjuck Dam

Burrinjuck Dam			
(26,210)	(26,196)	(26,182)	(30,214)
(26,208)	(26,194)	(28,212)	(30,212)
(26,206)	(26,192)	(28,210)	(30,210)
(26,204)	(26,190)	(28,208)	(30, 208)
(26,202)	(26,188)	(28,206)	
(26,200)	(26,186)	(28,204)	
(26,198)	(26,184)	(28,202)	

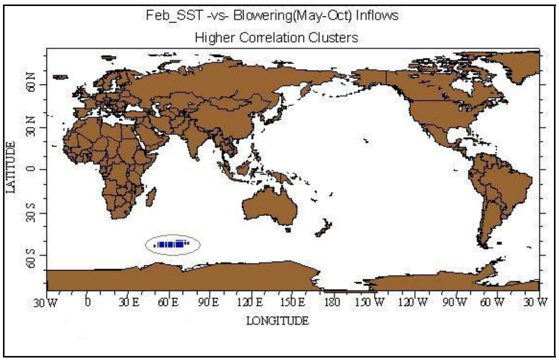

Figure 1. High correlated sea surface temperature clusters related to inflows to Blowering dam during May to October –The cluster is coloured in blue.

Southern Oscillation Index (SOI) was also used as an input to networks as SOI is a considerable climate factor. The input parameters of the models for the sole output 'January water allocation' are tabulated in Table 3 formulating relationships.

Each of the above data sets were organised in row and column-wise matrix for training data set. Some of the data rows were used as the cross validation dataset which were assigned to monitor any possible overtraining. Past January allocation data relevant to aforementioned training and cross validation data sets, was referred to networks as desired data set from which networks could calculate the error between network output and desired data. Since back propagation incorporated topologies were used, the error was propagated back through the network for error minimisation.

In the training processes there were 109 rows for R1 relationship extracted water allocation data from 1891 to 1999. For R2, R3 and R4 relationship, 53 rows of data from 1947 to 1999 period, were used in networks since the sea surface temperature and southern oscillation index data sets were reliable after 1947.

Table 3. Input parameters used in ANN models; √ indicates inputs assigned for the relationship**.**

Relationship	**August Allocation (AA)**	**January Allocation Risk Factor (JAR)**	**SST 1 (Blowering Dam Related SST)**	**SST 2 (Burrinjik Dam Related SST)**	**SOI**
R1: AA→JA	√	√			
R2: AASST→JA	√	√	√	√	
R3: AASOI→JA	√	√			√
R4: AASSTSOI→JA	√	√	√	√	√

i.e

- *R1*: JA = f(AA,JAR).
- *R2*: **JA = f(AA,JAR,SST1,SST2).**
- *R3*: **JA = f(AA,JAR,SOI).**
- *R4*: JA = (AA,JAR,SST1,SST2,SOI)

3.2. ANN Training

Training started with Generalised Feed Forward (GFF) networks that have back propagation-training rule incorporated into the basic topology called Multi Layer Perceptron (MLP). The GFF is powerful enough to generalise inputs and train networks to find possible relationships even in non-linear situations.

GFF networks were trained and optimised by changing the internal parameters. This provided a good relationship with better correlation coefficients. Similarly this procedure was applied for different topologies such as Jordan and Elman Networks (JEN), Radial Basis Function (RBF) and Self Organising Feature Maps (SOFM) etc.

Among the abovementioned topologies, the RBF was found to be the best topology that provided significant learning for the aforementioned relationships. The RBF has been constructed using the mathematical function in Equation 1 (Neurosolutions 4.32, Lefebvre et al., 2003) in a hidden layer with appropriate number of PEs. Inputs are directed from the input layer.

$$G(x; x_i) = \exp\left[\frac{-1}{2\sigma_i^2}\sum_{k=1}^{p}(x_k - x_{ik})^2\right]$$

Equation 1. Radial Basis Function – i^{th} node of the hidden layer 0; G - p multivariate Gaussian function; σ_i – variance of p data points, x_i – mean at i^{th} node

A network with RBF function is illustrated in Figure 2. The established networks were trained in 50000- 60000 iterations for many runs whilst changing network parameters.

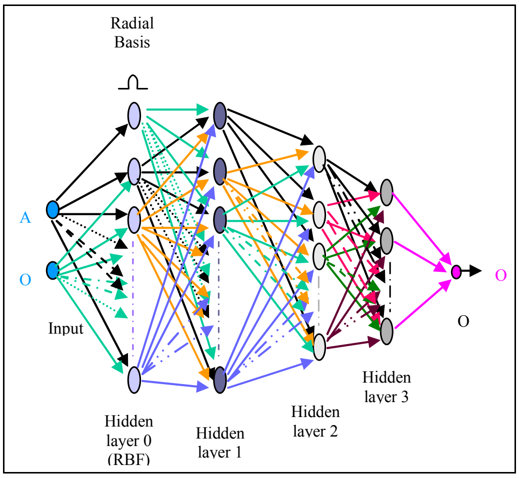

Figure 2. Network used to develop R1 relationship – Four hidden layers with 25, 25, 20 and 15 PEs in each layer respectively. Hidden layer 0 is with RBF functions.

3.3. Training Results

Figure 3 shows that the trained network responded progressively on test data set for ***R1*** relationship and similar results were found for rest of the relationships listed in Table 3. Relevant performance measures in column R2 of Table 4 indicate a higher value of correlation coefficient (r) about 0.99 between actual and network generated January water allocations.

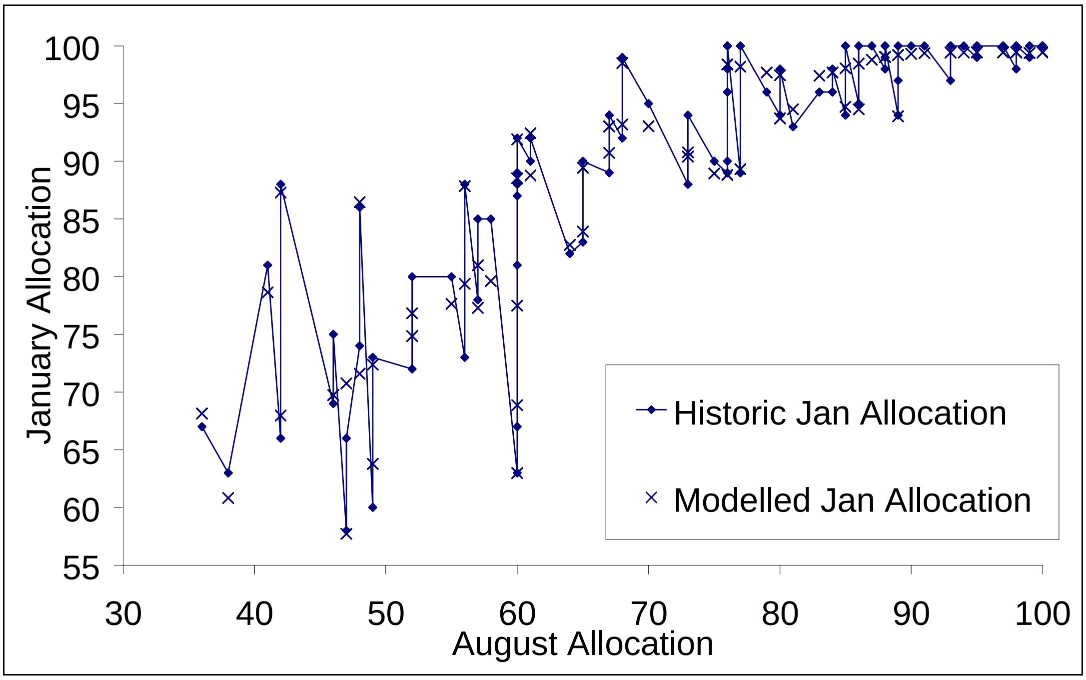

Figure 3. Actual and Network forecast January water allocation – plot against August water allocation (R1 relationship)

4. NUMERICAL VALIDATION OF ANN MODELS

Validation of the above ANN models was carried out using test data using the historic data to generate output. Comparison between actual and network output for test data for the ***R1*** relationship is shown in Figure 4, followed by the performance measurements in Table 5. Figure 4 shows that the network output and the actual January water allocation matched well; similar results were found for rest of three relationships listed in Table 3.

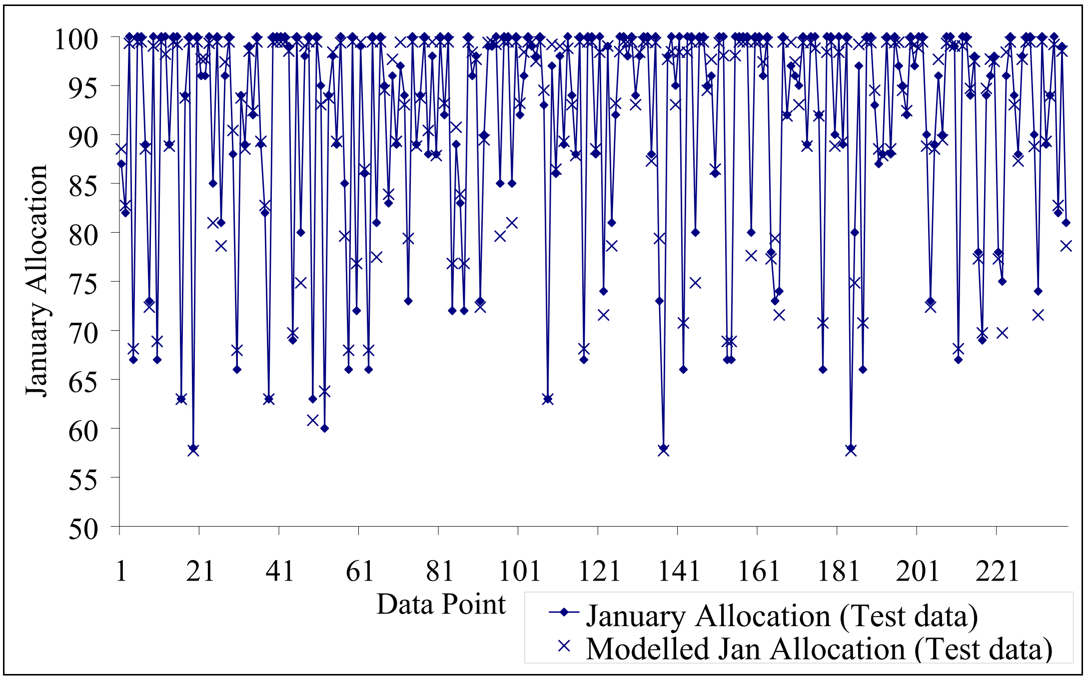

Figure 4. Training performance - Actual January allocation compared with Network output for test data set (R1 relationship)

In addition to the above plot, correlation coefficient above 0.99 and low values of mean square error((MSE), normalised mean square error(NMSE), and mean, minimum and maximum absolute errors, according to Table 5 support the reasonable validation of these models qualifying for forecasting.

.

Table 4. Performance measures for test data (for all relationships)

.

Performance Type	*Performance for January allocation of each relationship*			
	R1	***R2***	***R3***	***R4***
MSE	0.07569	0.01120	0.06209	0.02911
NMSE	1.03533	0.05410	0.56246	0.40143
Mean Abs Error	1.23379	0.31821	0.93520	0.77890
Min Abs Error	0.02552	0.00831	0.01378	0.00464
Max Abs Error	6.37968	1.95563	8.60058	3.20618
Correl. Coefficient	0.98771	0.99857	0.99043	0.99485

4.1. Sensitivity analysis

Sensitivity analysis could be used to ascertain the relative loading of each input channel contributed to the trained network. Calculation of partial derivatives of the output that incorporated with chain rule provides the output "y_k" respect to the input "x_i" provides the sensitivity component related to input x_i.is given by Equation 2. Table 5 shows the degree of each input contributed in the training of SST and SOI incorporated neural network.

$$S_{ik} = \frac{\partial y_k}{\partial x_i} = f'(net_k)\sum_{j=1}^{L} v_{jk} f'(net_j) w_{ij}$$

Equation 2. Sensitivity component of output k of input i

Where

S_{ik} or $\frac{\partial y_k}{\partial x_i}$ = Sensitivity of output y_k based on changes in input x_i

$f'(net_k)$ = derivatives of activation function of output neuron k

$f'(net_j)$ = derivatives of activation function of hidden neuron j

$w_{ij,}$ = weight between input x_i and hidden neuron j

v_{jk} = weight between hidden neuron j and output neuron k.

Table 5. Parameter contribution measurements at training of SST & SOI incorporated network

Parameter	*Sensitivity %*
August Allocation	40.9
January Risk Factor	37.3
SST Blowering	14.8
SST Burrinjuck	2.4
SOI	4.6

5. FIELD VALIDATION OF ANN MODELS

The model results were tested by predicting January water allocations for from 1999 to 2005 based on the August allocations for the last year and the trained ANN described in previous sections. Tables 6, 7 and 8 show overall performance of models R1, R2 and R3 in real situation.

Table 6. January Water Allocations Predictions Based on Historic Data Only – Model R1

Primary						
	Actual Water Allocation %		Model Predictions of January Allocation			
			Risk Factor			Difference
Year	August	January Next Year	25%	50%	75%	@ 50% risk
1999	50	73	50	61	67	-12.4
2000	59	90	66	82	83	-7.7
2001	47	72	47	54	62	-18.0
2002	38	38	55	55	57	16.9
2003	17	41	25	33	37	-7.8
2004	20	39	25	34	37	-5.4
2005	21		25	34	37	

Table 7. January Water Allocations Predictions Based on Historic Data and SST – Model R2

SST Incorporated						
	Actual Water Allocation %		Model Predictions of January Allocation			
			Risk Factor			Difference
Year	August	January Next Year	25%	50%	75%	@ 50% risk
1999	50	73	60	78	95	5.4
2000	59	90	63	79	93	-11.0
2001	47	72	55	76	96	3.8
2002	38	38	57	79	97	40.6
2003	17	41	44	78	93	37.4
2004	20	39	32	68	97	29.5
2005	21		41	75	96	

Table 8. January Water Allocations Predictions Based on Historic Data and SOI – Model R3

SOI Incorporated						
	Actual Water Allocation %		Model Predictions of January Allocation			
			Risk Factor			Difference
Year	August	January Next Year	25%	50%	75%	@ 50% risk
1999	50	73	50	65	84	-8.0
2000	59	90	62	90	87	-0.2
2001	47	72	47	57	82	-15.1
2002	38	38	38	45	76	7.2
2003	17	41	20	39	69	-1.7
2004	20	39	21	40	71	1.4
2005	21		21	40	71	

From the comparison of the above prediction results historic data and SOI based ANN model shows that at 50% risk factor the model results are very close to the actual announced January general security water allocations. SST incorporated model R2 overestimated January water allocations for the 2002-2004 period. This may be due to exceptionally low starting water allocations and borrowing of water from the future years which was outside the training data sets.

The model performance is being further improved by retraining the ANN with the latest data sets.

6. DISCUSSION

All the ANN models performed well with appropriate networks. Validity of each model has been found to be excellent as it shows very low error values and very high correlation coefficient within the range of 99-100%, between network output and actual values of January water allocation.

Building the **R1** model was started with two input parameters followed by **R2, R3 and R4** model development with changing inputs or increasing number of inputs. In terms of ANN technology all models learnt very well from the provided inputs.

Sensitivity analysis (Table 5) provided by the R4 case (other cases have less number of input parameters) shows the contribution of each input parameter to its network. However the sensitivity of input parameters could vary depending on the network in the same case or in different cases. For example although SOI has shown lowest contributions in R4 case it has shown higher values in different networks. R2 model has been observed to give lowest errors and highest correlation coefficient exists among R1, R2, R3 and R4 relationships as depicted in Table 3.

7. CONCLUSIONS

This study has shown that water allocations at the start of the irrigation season incorporating a farmer's risk factor could be utilised to forecast water allocation at the end of peak irrigation demand season. The SST incorporated model, SOI incorporated model and SST/SOI incorporated model have shown capability to forecast end of the irrigation demand season water allocation. The following conclusions are drawn from the results of this study:

- All the relationships that have been developed in this study, demonstrated ANN capability of forecasting end of January water allocation.
- SOI incorporated **R1** model is the most promising forecasting tool that shows good performance during the field testing of the model.
- The adaptive nature of ANN architecture is proven here as networks with appropriate additional input parameters learnt very well. .
- High performance indicators proved ANN capability of handling non-linearity.

8. REFERENCES

Khan, S., D. Robinson, R Beddek, B Wang ., D Dassanayake, T Rana (2004), Hydro-climatic and Economic Evaluation of Seasonal Climate Forecasts for Risk Based Irrigation Management, C*SIRO Land and Water Technical Report* No. 5/04, January, pp 10 – 43.

Lefebvre, C., J. Principe, D. Samson, D. Wooten, G. Geniesse, C. Fancourt, J Gerstenberger, N. Euliano, N Lynn, M Allen (2003), *Neurosolutions 4.32* , Neurosolutions, Inc.

Holger, R.M and G.C. Dandy (2000), Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 15, 101 – 124, Department Civil and Environmental Engineering, The University of Adelaide, S.A. 5005, Australia.

9. ACKNOWLEDGEMENTS

The authors wish to acknowledge funding support by the Cooperative Research Centre for Sustainable Rice Production and the Murray Irrigation Limited.

Water allocation data provided by the Department of Land and Water Conservation is appreciated.