Spatial Modeling of Electrical Conductivity with Neural Network

C. Sivapragasam*
Professor, Department of Civil Engineering, Kalasalingam University,
Krishnankoil, Srivilliputtur, Taminadu State, India – 626 190. Ph: 91-4563-289042 Ext. 310. Fax: 91-4563-289322. Email: sivapragasam25@gmail.com

V. Jegatheesan
Associate Professor, School of Engineering, James Cook University, Townsville,
QLD 4811, Australia. Ph: 61-7-4781-4871. Fax: 61-7-4781-6788. Email: jega.jegatheesan@jcu.edu.au.

V.M. Arun
Senior Research Fellow, Department of Civil Engineering, Kalasalingam University,
Email: arundas84@gmail.com

S. Vanitha
Lecturer, Department of Civil Engineering, Kalasalingam University,
Krishnankoil, Srivilliputtur, Taminadu State, India – 626 190. Ph: 91-4563-289042 Ext. 310. Fax: 91-4563-289322. Email: vnthsekar@yahoo.co.in

Abstract:
Many policy decisions for agricultural management in the coastal region closely depend on the extent of intrusion of sea water. In this study, Artificial Neural Network (ANN) is used to model the spatial variation of Electrical Conductivity to determine the extent of sea water intrusion in the coastal area of Brisbane, Australia. Quarterly EC data obtained from the observation (monitoring) wells located along the coast is used for training ANN architecture. The study demonstrates that ANN is able to model the spatial variation of EC with very good accuracy (even with very less training records) when some spatial information is used as one of the inputs in the network training. The results considerable improvement when compared with the network trained without the distance information.

Keywords: Spatial Modeling; Electrical Conductivity; Salt Water Intrusion; ANN.

1. Introduction
Saltwater intrusion refers to the replacement of fresh water in coastal aquifers by saltwater due to the motion of a saltwater body into the freshwater aquifer. Overexploitation of ground water by pumping is one of the major causes for salt water intrusion in the coastal regions. At present, many coastal aquifers in the world, especially shallow ones, experience an intensive saltwater intrusion caused by both natural and human-induced processes [Abd-Elhamid and Javadi, (2008), Bastani et al. (2008), Papadopoulou et al. (2005)]. Modeling the intrusion phenomena will greatly help in taking appropriate policy decisions for agricultural management.

Use of coastal aquifers as operational reservoirs in water resource systems requires the development of tools that facilitate the prediction of the aquifer behavior under different conditions. Quantitative understanding of the patterns of movement and mixing between freshwater and saltwater, as well as the factors that influence these processes, are necessary to manage the coastal groundwater resources (Ranjan, 2007).
Traditionally, salinity along the coastal regions has been measured using monitoring instruments (wells) installed at various locations [Gajendragad et al (1986), Fitterman and Deszcz-Pan (1999), Narayan et al. (2003)]. These measurements are used to predict future conditions and evaluate the water management alternatives. One of the indicators which is easy to measure and is also very informative as far as assessing the water quality for salinity is concerned is the Electrical Conductivity (EC). The presence of charged particles increases the conductivity of water when compared to pure liquid water which has a very low EC. Thus spatial modeling of EC in a region can give fairly good idea about the extent of sea water intrusion. EC measurements also useful for estimation of aquifer parameters such as hydraulic conductivity [Koukadaki et al. (2007)].

In this study, EC measured at the monitoring wells located near the coastal region of Brisbane (Australia) is modeled with Artificial Neural Network (ANN) to ascertain the extent of sea water intrusion. In the recent past many works have been reported in EC modeling with ANN due to its ability in modeling complex non-linear problems. For instance, Tutmez et al. (2006) attempted Neuro-fuzzy modeling of EC variation as a function of concentration of dissolved solids (such as sodium, potassium, calcium and magnesium) whose variation is space can be considered as non-linear. Similarly, Najah et al (2009) employed a complex 2 hidden layer neural network architecture for EC modeling in the study. Rajkumar and Thompson (2002) used Genetic Algorithm to search through combinations of ANN input parameters looking for the set of inputs that optimized the training criteria for temporal variation of salinity at hourly intervals for any time period in the year. Coppola et al (2005) developed ANN to predict the specific conductance values in unconfined coastal aquifers using input variables as initial conductance, total precipitation, mean daily temperature and total pumping extraction and established the potential of ANN to model such complex process. However, Kamp and Savenije (2006) found ANN to be not so effective when it is used as a coupling mechanism for an estuarine salt intrusion model. Tutmez et al. (2006) used ANFIS to model the non-linear relationship between EC and total dissolved solids.

As far as application of ANN in spatial modeling of EC is concerned, Farifteh et al. (2007) compared PLSR and ANN for soil EC variation. However, to the authors’ knowledge, no work has been reported yet for spatial modeling of EC with ANN in salt water intrusion problems. Since, EC variation is complex (due to consequence of a complex process), it is imperative that in order to obtain a better picture of spatial variation, EC values need to be suitably estimated in between the monitoring wells.

2. Study Area

Figure 1 shows the area where the data is collected for conductivity analysis. The data is collected by the Department of Natural Resources and Mining. There are two main areas susceptible to water logging, the Bruce Highway crossing of the Haughton River and at the coordinates 19.3 S, 147 E. The study area is divided primarily into three zones (Figure 2) based on the EC range with zone I containing some typically higher values of EC (more than 15000 mhos) in a given quarter of a year as compared to zone II (with less than 9000 mhos) during the same quarter. Sub-models, if necessary, can be formed within the zones for accuracy in modeling. This study focuses on zones I and II (Figure 2). Very high EC values are unacceptable for the long-term sustainability of the aquifer system.
In the entire study area, there are about 800 observation wells. It is observed that the measurement of EC is not done systematically in these wells. Consequently there are many missing data in the EC time series during the entire observation period. This is evident from Figure 3(a) and (b) from the number of data points on the graph. Prior to 2000, the data are obtained very randomly. After 2000, the data collection seems to be more systematic with at least one EC value in each quarter of the year for at least some of the wells. This posed the maximum difficulty in selecting appropriate number of wells for modeling. In the present study, only those wells are considered which contain EC values consistently for a given quarter (i.e. without any break in the EC time series).
In zone I, a total of 30 wells were found to contain EC information consistently for first quarter of 2000 and 2002, and are used for training and validation. Out of 30 records, 18 were used for training, 10 for testing and 2 for validating. Well no. 11900150 and 11900069 are used for validation primarily because of the following reasons:
The wells are located on the rightmost extreme of the study area. This makes the choice of neighboring wells difficult.
The measured EC value is very different for these two wells indicating a drastic spatial variation of EC in a short distance.

In zone II, a total of 50 wells are found to contain EC information for the first quarter of 2002. Out of this, 25 are used for training, 15 for testing and remaining 10 for validation. EC values range between 600 – 8500 mhos. Neural Network model is built for wells identified in zone I as well as zone II. The location of wells used for validation is shown in Figure 4.
3. Artificial Neural Network

Artificial Neural Network (ANN) is briefly introduced in this section. Multilayer feed forward network with Back Propagation learning algorithm is one of the most popular neural network architectures, which has been deeply studied and widely used in many fields. Typically a neural network consists of three layers: (1) an input layer; (2) an output layer; and (3) an intermediate or hidden layer. The input vectors are $D \in \mathbb{R}^n$, and $D = (X_1, X_2, \ldots, X_n)^T$; the outputs of $q$ neurons in the hidden layer, $Z \in \mathbb{R}^q$, and $Z = (Z_1, Z_2, \ldots, Z_q)^T$; and the outputs of the output layer are $Y \in \mathbb{R}^m$ and $Y = (Y_1, Y_2, \ldots, Y_m)^T$: Assuming that the weight and the threshold between the input layer and the hidden layer are $w_{ij}$ and $\theta_j$, respectively, and the weight and the threshold between the hidden layer and output layer are $w_{jk}$ and $\theta_k$, respectively, the outputs of each neuron in a hidden layer and output layer are:

$$Z_j = f \left( \sum_{i=1}^{n} w_{ji} X_i - \theta_j \right)$$  \hspace{1cm} (1)

$$Y_k = f \left( \sum_{j=1}^{q} w_{kj} Z_j - \theta_k \right)$$  \hspace{1cm} (2)

where $f(\cdot)$ is a transfer function, which is the rule for mapping the neuron's summed input to its output, and by a suitable choice, is a means of introducing a non-linearity into the network design. One of the most commonly used functions is the sigmoid function, and it is monotonic increasing and ranges from 0 to 1. $\theta$ is the bias. Due to non-linear nature of the problem being studied, different types of sigmoid functions such as the hyperbolic tangent function are used in ANN training. Details on ANN can be found in Karunanithi et al. (1994). In this study Neuroshell ®2, Release 4.0 software is used for implementing ANN.

4. Model Development

The most important task is to choose appropriate input variables as well as the architecture for the ANN model. The EC value at a given point will obviously be more closely related with the EC values in surrounding wells (neighboring wells) than those which are far off. After a closer analysis of the proximity of different wells to

Figure 4: Location of wells chosen for validation in zone I and zone II
each other, it is decided to use 3 neighboring monitoring wells which lie within a distance of about 3 km from
the location where estimation of EC is desired. This is the maximum diameter of spatial EC estimation. Then,
distance to these wells is estimated from an arbitrary origin decided based on the available easting and northing
information of the wells. For this study, this origin is taken at (500000m, 7795000m) as shown in Figure 2.

In this study, two models are studied viz., ANN-I and ANN-II which are described in detail as below.

4.1.Model: ANN-I
Since 3 neighboring monitoring wells are selected to model spatial variation of EC, the input to ANN-I
consisted of 3 variables viz., the EC value of each of the three neighboring monitoring wells. The input
variables can be symbolically represented as:

\[ I = f(EC_1, EC_2, EC_3) \] (3)

Where,
EC_1 = EC value of the first monitoring well in the proximity.
EC_2 = EC value of the second monitoring well in the proximity.
EC_3 = EC value of the third monitoring well in the proximity.

ANN architecture is optimized after many trials as a 3-layer feed forward network with input layer consisting of
3 nodes (each of them corresponding to the input variables as discussed above), hidden layer with 9 nodes and
the single output. The logistic curve is used as activation function or transfer function which is used by the
Neuroshell ®2, Release 4.0 software as one of the non-linear mapping function (similar to hyperbolic tangent or
sigmoid). Learning and momentum rate converged to an optimal value of 0.6 after a few trials. During ANN
training, the training and testing error systematically reduced with the progress in number of epochs. The
training was stopped at best test error. Since the training records are too few, it was ensured that the network is
not over trained. This is evident from the training and testing error measures. For example, the correlation
coefficient for training, testing and validation data used in zone II are 0.92, 0.83 and 0.82 respectively. Table 1
and 2 show the validation results for the wells in zone I and II respectively.

As it can be observed from Table 1, ANN-I does very poorly in predicting for both the wells with absolute error
being 10535 mhos and 6387 mhos respectively for the year 2000. However, for year 2002, the EC value is
predicted with an absolute error of 76 mho for well no. 11900069, while well no. 11900150 is very poorly
estimated.

Similarly, from Table 2, the ANN-I does poor prediction in almost all the 10 validation data. In some cases
(well no 12000812 and 12000826), the predictions are highly over estimated and under estimated respectively.

4.2.Model: ANN-II
A modification is done in the inputs to ANN. In geostatistical terms, the distribution of difference in values (in
this case EC values) between two positions depends on distance between them and their relative orientation. A
similar approach is adopted here with a slight modification that instead of distance between the neighboring
wells, the distance to the neighboring wells from a common origin is used as additional input in ANN training.
Consequently, the input pattern is symbolically represented as:

\[ I = f(EC_1, D_1, EC_2, D_2, EC_3, D_3, D_3os) \] (4)

Where,
EC_1 = EC value of the first monitoring well in the proximity.
D_1 = Distance to the first monitoring well from the assumed origin.
EC_2 = EC value of the second monitoring well in the proximity.
D_2 = Distance to the second monitoring well from the assumed origin.
EC_3 = EC value of the third monitoring well in the proximity.
Here again, the ANN architecture is optimized after many trials as a 3-layer feed forward network with input layer consisting of 7 nodes (each of them corresponding to the input variables as discussed above), hidden layer with 12 nodes and the single output. The Gaussian function is used as activation function. Table 1 and 2 show the validation results.

Table 1: ANN-I and ANN-II predictions for 2000 and 2002 for Zone I

<table>
<thead>
<tr>
<th>Well no.</th>
<th>2000</th>
<th>2002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>measured EC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(mhos)</td>
<td></td>
</tr>
<tr>
<td>measured EC</td>
<td>18 140</td>
<td>704</td>
</tr>
<tr>
<td>119000150</td>
<td>7605</td>
<td>7091</td>
</tr>
<tr>
<td>11900069</td>
<td>14 359</td>
<td>1720</td>
</tr>
<tr>
<td>ANN-I</td>
<td>11900150</td>
<td>11900069</td>
</tr>
<tr>
<td></td>
<td>790</td>
<td>1799</td>
</tr>
<tr>
<td>ANN-II</td>
<td>11900150</td>
<td>11900069</td>
</tr>
</tbody>
</table>

As seen from the table 1, except for well no. 11900069 during year 2002, ANN-II performs relatively much better. Particularly of interest is to note the accuracy with which the high EC value of 18140 and 16640 mhos is estimated given the fact that the ANN model is trained with very limited records. The effect of including the additional ‘distance’ variable can be clearly seen by means of the improvement in prediction results.

For the wells in zone II (Table 2), ANN-II clearly improves the prediction in 5 of the wells namely 12000873, 11910968, 12000826, 12000940, and 12000190 when compared with ANN-I. For well no. 12000812, where ANN-I estimates 7783 mhos against the measured value of 818 mhos, ANN-II brings down this high overestimation to 3520 mhos. Though this is still a high overestimation, it is interesting to note the effect of including ‘distance’ information in the input in improving the prediction. Similarly, the underestimation by ANN-I for well no. 12000826 against the measured value of 8070 mhos is improved by ANN-II to 6848 mhos which is more closer to the measured EC.

A closer look at the estimated EC values (Table 1) shows that the model (ANN-I and ANN-II) estimations for well no. 11900150 are much lower than the measured EC values, meaning the both models produced under estimation for 2000 and 2002. While estimated values of these models (ANN-I and ANN-II) during 2000 and 2002 for well no. 11900069 are higher than the measured EC in these periods. This is probably because the ANN training is stopped based on best average test results, and due to lesser number of training records, the model finds difficult to arrive at stable weight vector. Consequently the model prediction on two largely different EC values shows contradictory behavior.
Table 2: ANN-I and ANN-II predictions for 2002 for Zone II

<table>
<thead>
<tr>
<th>Well No.</th>
<th>Measured EC (mhos)</th>
<th>ANN-I</th>
<th>ANN-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>12000873</td>
<td>1275</td>
<td>856</td>
<td>1024</td>
</tr>
<tr>
<td>12000812</td>
<td>818</td>
<td>7783</td>
<td>3520</td>
</tr>
<tr>
<td>11910975</td>
<td>1256</td>
<td>868</td>
<td>620</td>
</tr>
<tr>
<td>11910968</td>
<td>776</td>
<td>924</td>
<td>704</td>
</tr>
<tr>
<td>12000190</td>
<td>1536</td>
<td>858</td>
<td>781</td>
</tr>
<tr>
<td>11910934</td>
<td>1076</td>
<td>890</td>
<td>660</td>
</tr>
<tr>
<td>12000826</td>
<td>8070</td>
<td>5946</td>
<td>6848</td>
</tr>
<tr>
<td>12000940</td>
<td>641</td>
<td>882</td>
<td>634</td>
</tr>
<tr>
<td>11910860</td>
<td>1536</td>
<td>1846</td>
<td>3927</td>
</tr>
<tr>
<td>12000190</td>
<td>900</td>
<td>1700</td>
<td>1027</td>
</tr>
</tbody>
</table>

5. Conclusion
In this study ANN is used for modeling the spatial variation of EC in the Brisbane coastal area. Based on the study, the following conclusions can be arrived at.

The inclusion of some spatial information (in this study distance to the wells from assumed origin) seems to be a suitable input in the ANN modeling as far as spatial variation of EC is concerned.

Though, due to limitation in the number of training records, ANN-II results are not able to show considerable improvement in the estimation of EC in some cases, there is clearly an improvement when compared with ANN-I which do not have the ‘distance’ information.

The appropriate choice of input variables, where possible, can at least to some degree, help overcome the need for lengthy training records usually required for ANN training.

Acknowledgement
Dr. Kumar Narayan of Davis Laboratory, CSIRO, Townsville and the Department of Natural Resources and Mining, Queensland have been acknowledged greatly – the former for helping the second author to formulate and model the salt water intrusion in the Burdekin River Irrigation Area (BRIA) and the later for providing electrical conductivity data. The first author wish to acknowledge the financial support and the computing facilities extended by the Council of Scientific and Industrial Research (CSIR), Government of India, through their sponsored research scheme. The authors also wish to extend their sincere thanks to the anonymous reviewers who have given very meaningful and insightful comments to bring the manuscript to the present form.

Author Disclosure Statement:
No competing financial interests exist.

References