SALINITY RISK PREDICTION USING LANDSAT TM AND DEM-DERIVED DATA

F. H. Evans, P. A. Caccetta

CSIRO Mathematical and Information Sciences
Private Bag No. 5 PO Wembley 6913
Western Australia
Phone: 08 9333 6177 Fax: 08 9333 6121
Email: Fiona.Evans@cmis.csiro.au

Abstract
This paper presents a method for predicting areas at risk from dryland salinity using information derived from multi-temporal Landsat TM satellite images combined with landform data derived from high-quality digital elevation models. The method is applied in the south west agricultural region of Western Australia to predict areas at risk from dryland salinity. This paper presents modifications to previous methods suggested by the use of high-quality elevation data previously unavailable in WA.

The method aims to reproduce expert opinion about the future extent of salinity by using decision trees to determine the relationship between salinity risk and variables that describe various aspects of the landscape. Feature selection procedures are used to determine the optimal subset of variables for predicting risk areas. Preliminary studies were conducted in five subcatchments and the model extrapolated over 30 000 km$^2$ to produce maps of those areas expected to become saline under current management practices.

1 Introduction
The Land Monitor project aims to provide information about land condition, specifically salinity and the status of remnant vegetation, for the whole of the south-west agricultural region of Western Australia. It is a collaborative project involving Agriculture WA, CSIRO, Conservation and Land Management, the Department of Land Administration, Water and Rivers Commission, the Department of Environmental Protection and Main Roads WA.

Satellite images, digital elevation models and ground data will be used to produce maps of land condition, such as the current extent of salinity, the changes since 1990 and predictions of future salinity risk. The methods used for predicting future salinity risk were developed as part of a project funded by the Land and Water Resources Research and Development Corporation as part of the National Dryland Salinity Project (see Evans et al., 1995).

The project investigated methods for predicting areas at risk from salinity. It focussed on methods that aimed to emulate expert-knowledge in the form of rule-based systems. A workshop held in Bunbury on June 20 - 21 in 1994 provided a forum to quantify current...
knowledge in explicit rule-based form. The following factors were identified as indicators of salinity risk:

- time since clearing
- depth to ground water and rate of ground water rise
- distance to existing salinity and degree of waterlogging
- climate
- geology and depth to basement
- soil type and available salt storage
- landform type: including factors like convergence, relative height, drainage density, drainage slope, valley size, position in the flow path
- historical and existing vegetation cover
- hydrogeology: vertical structures, including shears, faults, dykes, and bedrock highs and horizontal structures, including regolith stratigraphy
- land management

The project obtained all available data relating to these factors. Landsat TM satellite data were used to produce maps showing areas currently affected by salinity and areas supporting remnant vegetation. Digital elevation models (DEMs) were used to produce landform descriptors. Agriculture WA provided digital maps showing interpreted landform systems, soil systems, geology and hydrogeological structures. The datasets were systematically evaluated for their ability to predict salinity risk and the final subset of input data were reduced to Landsat TM and DEM-derived variables.

The project showed that using a decision tree classifier, salinity risk areas as defined by an expert could be predicted with 78% accuracy whilst non-risk areas were predicted with 80.5% accuracy. The predicted risk map was evaluated in the field by AgWA hydrologists with assistance from local landholders and the on-ground evaluation supported the credibility of the results.

This paper describes modifications to the method following the acquisition of accurate DEMs as part of the Land Monitor project. The new DEMs have been produced from 1:40 000 scale photography using automated softcopy techniques that collect heights on a dense grid across the parts of the south west agricultural region of WA that previously had available only 5, 10 or 20m contour densities. The DEMs are expressed on a 10m grid and have accuracies of 1-2m in elevation. Their availability enables the derivation of accurate spatial data describing landform attributes that relate to the location and rate of change of salt-affected land.

The results described in this report have been produced using 50% of the available ground data for training and the remaining 50% for testing. The classifier c5.0 (Quinlan, 2000) has been used to train the decision trees with options -c0 and -m5 and costs of 2:1 for errors of omission versus errors of commission. The results for individual study areas have been produced by training and testing only on local data (ie. data from that study area). Regional
results are also presented and the method for salinity risk prediction is applied to predict risk in the Dumbleyung and Mt Barker regions (shown in Figure 1).

Throughout this paper, the term **predicted risk areas** is assumed to mean areas that are predicted to have high water-tables in the future (i.e. discharge areas), parts of which are likely to become saline. Whilst waterlogging may be common, predicted risk areas will not necessarily be salt-affected throughout.

![Figure 1 The study area includes Dumbleyung in the north and Mt Barker in the south.](image)

2 **Salinity prediction methodology**

The steps used to produce the salinity risk maps are listed below. This is the standard methodology being used for Land Monitor salinity prediction.

1. Obtain ground data in the form of areas predicted to be at risk from salinity by expert hydrologists, delineated on airborne photographs or hardcopy maps.

2. Use the Land Monitor DEM and current salinity maps to create the following derived variables:
   - average upslope height
   - average upslope slope
   - flow slope (or average downhill slope)
   - height above nearest salt within a watershed
The methods used to derive the above variables are described in Caccetta (1999).

3. Create a decision tree that relates the derived variables to salinity risk for the ground truth areas provided (Evans et al., 1996).

4. Use the tree to extrapolate the decision tree model and predict risk areas.

Modifications to the method are required to determine the spatial extent to which the decision tree model can be extrapolated. Regional differences may mean that a decision tree trained in one region may not extrapolate well to another region. This is discussed in section 5 where both local and regional results are presented.

3 Ground data

Ground data has been provided by Agriculture WA hydrologists. The ground data were digitised from interpretations of areas at risk from salinity based upon the following data (where available):

- rate and extent of change of salinity
- existing data on groundwater levels and the rate at which groundwater is rising
- water balance calculations (eg. AgET) to estimate the difference between current recharge, discharge from saline
- areas and base flow
- interpreted hydrogeological data (shears, faults, dykes, bedrock highs)
- geology
- amount and type of remnant vegetation in the catchment and clearing history
- electromagnetic, radiometric and magnetic data
- ground water modelling where appropriate
- landscape processes (eg. areas with poor drainage and convergent inflow, concave inflection points (breaks of slope), non-saline discharge areas)

Ground data were provided for the following areas: Broomehill, Toolibin, Tambellup, Kent and the South Stirlings. Of these areas, Broomehill and Toolibin are located in the Dumbleyung region and the remainder in the Mt Barker region. Figure 2 shows the ground truth map for the Toolibin catchment.
Figure 2 Ground truth map for the Toolibin catchment. Risk areas are shown in blue.

4 Feature selection

Feature selection procedures are used to determine whether salinity risk can be accurately reproduced using a subset of the DEM-derived variables listed in section 2. By reducing the number of variables, we aim to reduce the size of the decision tree (rendering it simpler to interpret) and improve its accuracy and ability to extrapolate.

Exhaustive feature selection involves testing each possible subset of the variables and assessing the subsets according to their accuracy on a set of independent test data. The primary disadvantage of the exhaustive approach is the large number of subsets that must be tested (in this case, $2^9$). A number of heuristic approaches are available, of which we present the result of using simple forward selection. Preliminary studies were conducted using data for the Broomehill study area and assessed in the Toolibin, Kent, and Tambellup areas.

Table 1 Simple forward selection results in the Broomehill study area.

<table>
<thead>
<tr>
<th>number of variables</th>
<th>variable(s)</th>
<th>salt accuracy</th>
<th>not salt accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Height above Salt</td>
<td>60.4</td>
<td>90.6</td>
</tr>
<tr>
<td>2</td>
<td>+ Upslope Cleared Area</td>
<td>65.0</td>
<td>94.3</td>
</tr>
<tr>
<td>3</td>
<td>+ Height above Streams</td>
<td>68.5</td>
<td>93.2</td>
</tr>
<tr>
<td>4</td>
<td>+ Average Height of Upslope Area</td>
<td>68.6</td>
<td>92.9</td>
</tr>
<tr>
<td>4</td>
<td>+ Upslope Area</td>
<td>69.1</td>
<td>92.9</td>
</tr>
<tr>
<td>5</td>
<td>+ Flowpath Length</td>
<td>69.8</td>
<td>92.2</td>
</tr>
</tbody>
</table>
The best subset of five variables is: height above salt, upslope cleared area, height above streams, upslope area (water accumulation) and flow path length.

Tables 2-4 show the prediction accuracies for the Toolibin, Tambellup and Kent River study areas.

**Table 2 Toolibin results.**

<table>
<thead>
<tr>
<th>number of variables</th>
<th>salt accuracy</th>
<th>not salt accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>86.8</td>
<td>93.5</td>
</tr>
<tr>
<td>5</td>
<td>86.5</td>
<td>92.7</td>
</tr>
</tbody>
</table>

**Table 3 Tambellup results.**

<table>
<thead>
<tr>
<th>number of variables</th>
<th>salt accuracy</th>
<th>not salt accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>63.9</td>
<td>88.4</td>
</tr>
<tr>
<td>5</td>
<td>69.4</td>
<td>87.5</td>
</tr>
</tbody>
</table>

**Table 4 Kent results.**

<table>
<thead>
<tr>
<th>number of variables</th>
<th>salt accuracy</th>
<th>not salt accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>79.2</td>
<td>80.2</td>
</tr>
<tr>
<td>5</td>
<td>80.1</td>
<td>77.7</td>
</tr>
</tbody>
</table>

Note that the use of the 5 variables selected using the simple forward selection approach provides results that are similar or improved upon those achieved when all of the variables are used.

Tables 5 and 6 show the results for the Dumbleyung and Mt Barker regions. The five variables provide reasonable information for predicting salinity risk over the broader regions.

**Table 5 Dumbleyung results.**

**Table 6 Mt Barker results.**

<table>
<thead>
<tr>
<th>number of variables</th>
<th>salt accuracy</th>
<th>not salt accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>79.2</td>
<td>94.4</td>
</tr>
<tr>
<td>5</td>
<td>80.7</td>
<td>94.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>number of variables</th>
<th>salt accuracy</th>
<th>not salt accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>71.7</td>
<td>83.7</td>
</tr>
<tr>
<td>5</td>
<td>71.3</td>
<td>84.0</td>
</tr>
</tbody>
</table>

The following section discusses the extrapolation of results to the broader Dumbleyung and Mt Barker regions of south west WA.

**5 Risk prediction extrapolation**

This section describes the extrapolation of salinity risk prediction to the Dumbleyung and Mt Barker regions. The geographic region extends from the town of Corrigin in the north to the
south coast. It is divided into two broad hydrogeographical regions: the Dumbleyung region includes the upper Blackwood Basin and the Mt Barker region includes the Frankland-Gordon catchment, the Kent catchment and the Stirling Range National Park.

The following results are a sub-sample of experiments that were performed to ensure that the decision tree prepared would produce results that could be extrapolated to map risk areas for the Dumbleyung region. All of the following risk maps and statistics were produced using the decision tree classifier c5.0 with options -c0 -m5, and costs set such that errors for risk areas incur double the cost of errors in non-risk areas. The use of costs is required because there are significantly more non-risk training sites than risk training sites.

Table 7 shows the results obtained from using all the variables and the subset of variables chosen in section 4. In the table, global fitting used the combined Toolibin and Broomehill training data, while local fitting used only training data from the tabled catchment.

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall</th>
<th>Toolibin</th>
<th>Broomehill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>salt</td>
<td>not salt</td>
<td>salt</td>
</tr>
<tr>
<td>global fitting - 9 variables</td>
<td>79.2</td>
<td>94.4</td>
<td>87.7</td>
</tr>
<tr>
<td>local fitting - 9 variables</td>
<td>-</td>
<td>-</td>
<td>86.8</td>
</tr>
<tr>
<td>Toolibin training data only</td>
<td>-</td>
<td>-</td>
<td>86.8</td>
</tr>
<tr>
<td>Broomehill training data only</td>
<td>88.4</td>
<td>88.6</td>
<td>91.5</td>
</tr>
<tr>
<td>global fitting - 5 variables</td>
<td>80.7</td>
<td>94.2</td>
<td>88.7</td>
</tr>
<tr>
<td>local fitting - 5 variables</td>
<td>-</td>
<td>-</td>
<td>86.5</td>
</tr>
</tbody>
</table>

The highest accuracy for predicting risk areas in both Broomehill and Toolibin occurred when the Broomehill ground truth data were used to train the decision tree classifier. The classifier was used to produce a prediction map for the Dumbleyung region. Whilst mapping risk areas with sufficient accuracy, the prediction map tends to over-estimate the risk areas (by 7.1% in Broomehill and 12.5% in Toolibin). This problem can be reduced by combining the Broomehill and Toolibin ground truth data, but the accuracy with which risk areas can be mapped is then reduced (from 70.2% to 67.0% in Broomehill and from 91.5% to 87.7% in Toolibin). While the reduction in accuracy is small, visual inspection of the maps produced in this manner shows many streamlines predicted to go saline are missed.

The results also showed that while the Broomehill ground data can be extrapolated to accurately map risk areas in Toolibin, a decision tree trained using the Toolibin ground data will only map 53.9% of the risk areas in Broomehill. The predicted salinity risk map for the Toolibin catchment is shown in Figure 3.
Figure 3 Toolibin risk map overlaid on 1993 Landsat TM Band 4 in greyscale.

Table 8 shows the results of similar analyses performed for the Mt Barker region. The global decision tree has been used to produce a map showing salinity risk areas for the Mt Barker region.

Table 8 Mt Barker salinity risk prediction - summary of analyses.

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall</th>
<th>Tambellup</th>
<th>Kent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>salt</td>
<td>not salt</td>
<td>salt</td>
</tr>
<tr>
<td>global fitting</td>
<td>76.1</td>
<td>81.1</td>
<td>76.7</td>
</tr>
<tr>
<td>local fitting</td>
<td>-</td>
<td>-</td>
<td>78.3</td>
</tr>
</tbody>
</table>

Examination of the extrapolated risk maps south of this point suggests that risk areas are also being over-estimated in the higher rainfall zones and therefore additional training data are required for these areas.

7 On-ground validation

The prediction map for the Dumbleyung region has been produced using the decision tree classifier trained on only the Broomehill data with accuracies presented in Table 7. The accuracies achieved in the two areas with ground data were acceptable and further ground-truthing has been performed in the Boscabel and Towerinning areas. Detailed assessments of the salinity risk maps at specific geographically located areas were noted for both the Boscabel and Woodanilling areas. In general:

1. The overall impression of the risk maps is that they are good.
2. Errors in local valleys are of two types. Some should show less risk extending up the valley and some should show more.

3. Broad valley floors are exclusively mapped as risk. However, there are many areas where the groundwater is near-surface (ie. 1-2m) within the broad valleys that may be slightly higher in elevation than the true valley floor (less than 1m height differences can be significant) that will not become salt-affected. These areas may currently be supporting a healthy crop or pasture cover.

The prediction map for the Mt Barker region has been produced using the decision tree classifier with accuracies presented in Table 8.

Whilst the results are acceptable in the Kent, Cranbrook and Tambellup catchments, risk areas are severely over-estimated in the South Stirlings and southern higher rainfall areas. More training data are required to correct these problems.

8 Conclusions
This paper has presented a method for predicting areas at risk from dryland using salinity maps derived from multi-temporal Landsat TM satellite images combined with landform data derived from high-quality digital elevation models. The method uses the decision tree classifier c5.0 to emulate and extrapolate expert knowledge about salinity risk and produce spatial maps of the areas at risk in the Dumbleyung and Mt Barker regions of WA.

Improvements to the method since its development in the Kent catchment result from the use of high-quality DEMs to produce derived variables, and an improvement in the classification tools.

9 References
Caccetta, P. C. (1999), Some methods for deriving variables from digital elevation models for the purpose of analysis, partitioning of terrain and providing decision support for what-if scenarios, CMIS internal task report No. 99/164.

Evans, F. H. (2000), Feature selection for salinity risk prediction, CMIS internal task report, Land Monitor project.

Evans, F. H., Caccetta, P. C., Ferdowsian, R., Kiiveri, H. T. and Campbell, N. A. (1995), Predicting salinity in the Upper Kent River catchment, A report from the LWRRDC project Integrating remotely sensed data with other spatial data sets to predict areas at risk from salinity.


Acknowledgements
The authors greatly acknowledge the assistance and advice we have received at all stages from our collaborators. In particular, we thank Ruhi Ferdowsian, Arjen Ryder, Richard George and the other hydrologists in the Catchment Hydrology Group of Agriculture Western Australia. The Land Monitor project is partly funded by the National Heritage Trust and is an initiative of the WA State Salinity Action Plan.