

**Image fusion with conditional probability networks
for monitoring the salinisation of farmland**

by

Harri Kiiveri

CSIRO Mathematical and Information Sciences

The Leeuwin Centre, 65 Brockway Rd, Floreat, WA 6104.

email: harri.kiiveri@cmis.csiro.au

Phone: 08 9333 6315

Fax: 08 9333 6121

Peter Caccetta

CSIRO Mathematical and Information Sciences

The Leeuwin Centre, 65 Brockway Rd, Floreat, WA 6104.

email: peter.caccetta@cmis.csiro.au

RUNNING TITLE: Image fusion with conditional probability networks

KEY WORDS: satellite imagery, classification, uncertainty, data integration, conditional probability networks, spatial-temporal models, spatial context

ABSTRACT: We show how a series of satellite images used in conjunction with data derived from a digital terrain model can be used to monitor salinity in farmland. Using these data, a conditional probability network (CPN) is constructed to produce salinity maps. The maps are the result of combining uncertain information in images with uncertain knowledge or rules, where the rules are of a temporal and spatial nature. A specific model for extending conditional probability networks to handle the case of spatial context is given. To implement this model requires minor modifications to existing code for handling non-spatial CPN's.

1. Background and description of data

As a result of the clearing of native vegetation, large tracts of farmland in Western Australia have experienced increasing levels of salinisation, rendering significant areas of land unproductive. As part of an overall strategy to arrest the increasing salinisation of land, we have been involved in developing methods for monitoring the extent of salt-affected land using sequences of satellite imagery and landform data derived from digital terrain models.

This paper describes a spatial-temporal model which we have developed, which allows sequences of satellite images to be integrated with other spatial data sets to accurately map the extent of salt-affected land over time. The model is a conditional probability network (CPN), see [1]. It enables uncertainties in the available data to be propagated through the calculations and to be represented in the output products. The model presented here also uses spatial context within the images.

The methodology is illustrated with an example from the Ryans Brook catchment in the South West of Western Australia. The area is contained in a 1045 by 1045 twenty-five metre pixel image. A sequence of spring Landsat Thematic Mapper (TM) imagery from 1989, 1990, 1993 and 1994 was used. For each of these years, a maximum likelihood classification was performed to produce four classified images, each with six classes: 1 - salt, 2 - mixed salt/bare, 3 - bare, 4 - bush, 5 - agriculture, and 6 - water. These classified images were used as inputs to the model.

The TM imagery alone cannot distinguish between some salt-affected land and bare ground. Hence a temporal sequence of images and landform were used to assist in the mapping and monitoring. Effectively, we use the knowledge that salinity persists from year to year, and the notion that salt-affected land is much more likely to be found in valleys than on hilltops or upper slopes, to improve the mapping results.

Relative uncertainties in the classes of the classified satellite images were estimated by the posterior probabilities of class membership derived from the maximum likelihood classification. Note that a

probability of one of belonging to a particular class does not imply that a class label is correct but simply that the ground cover is spectrally similar to a specific training class and not to any of the other training classes. The landform map was derived from a digital terrain model using the methods described in [2]. Basically, landform categories were obtained by stratifying a water accumulation map into two broad classes called hilltops and valleys. Since the cutoff used to discretise the water accumulation map was somewhat arbitrary, uncertainty was introduced at the class boundaries. This was done by creating two probability surfaces, one for each class, with the probability of a particular class occurring at a specific location depending on the distance of its water accumulation value from the cutoff.

The uncertainties in the classification labels and landform boundaries were explicitly propagated through subsequent calculations. For more details see [2], [3] and [4].

2. The conditional probability network model

The model we developed can be represented graphically as in Figure 1 below. In the figure, the boxes denote observed images or maps while the circles denote unobserved images or maps. The classified images are denoted as C89, C90, C93 and C94 respectively, while the landform image is denoted by L. We wish to predict the unobserved true ground cover images T89, T90, T93 and T94, given the observed images.

The selection of an appropriate CPN graph is an important issue. For this example, we chose an intuitively reasonable structure. Interviewing experts in the subject matter, fitting models to a large sparse contingency table [5] or a combination of both are other methods for choosing a graph. Yet another alternative is to choose a model by minimising prediction error on a validation set of ground-truth data.

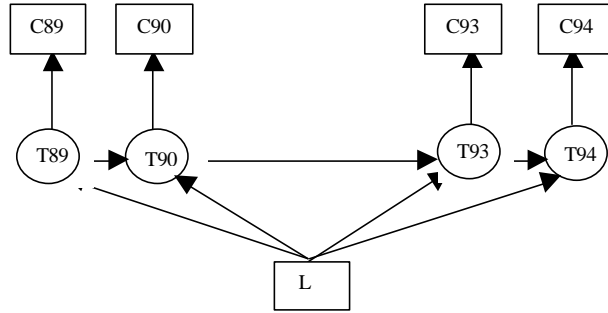


Figure 1: Graph of conditional probability network for monitoring salinity over time

Intuitively, the graph can be thought of as specifying parent-child relationships amongst the nodes of the graph, or equivalently, the images they represent. The parents of a particular child are the set of nodes with arrows pointing to that child. The arrows convey the notion that the parents directly "influence" their children. In Figure 1 we see that landform influences the true images which in turn influence the classifications derived from the satellite images. We also note that the true image at a particular time is influenced by the landform image and the immediately preceding true image. Influence is defined by a table of probabilities associated with each node (probabilistic rules). These tables define the probability of a class label at a particular point in an image given the class labels of its parent images at the same point and possibly at neighbouring points (see below). Note that the graph completely specifies the rules required by the system; however, the probabilities for the rules need to be supplied. There can also be assumptions of spatial correlation or independence amongst pixels in an image. For more details about these types of models, see Section 4, [1] and [6].

In terms of conditional independence constraints, there are other models which are equivalent to the model corresponding to Figure 1. For example, omitting the arrows, or reversing the arrows between the circles. We have chosen a specific interpretation from a consideration of the context of the problem.

For each location on the ground, the model in Figure 1 specifies a joint probability distribution for the nine image layers. This is a distribution for approximately 3.3 million (2×6^8) possible states and is specified by

the order of **300** independent probabilistic rules i.e. probability statements about the values of an image at a particular location given the values of its parent images at that location. For this particular application, a number of rules were specified from a priori knowledge of the system. Probabilities for the remaining rules were estimated from the data by using the EM algorithm [7], [8].

3. Some technical details

We include some technical details here about the methods used in this paper. To begin, we need to define some notation. We write $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4$ for the images C89, C90, C93, C94 ; $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ for the images T89, T90, T93, T94 and \mathbf{z} for the landform image L respectively. For example $\mathbf{y}_1 = (y_{i1}, i = 1, \dots, n)$ and $\mathbf{x}_1 = (x_{i1}, i = 1, \dots, n)$ where n denotes the number of pixels in each image (assumed the same) and x_{is}, y_{is} etc denotes the label at pixel i for image s . Probability distributions will be denoted as $p(\cdot)$ e.g. $p(x_{i1} x_{i2} \dots x_{in})$ and conditional distributions by $p(\cdot | \cdot)$ e.g. $p(y_{i1} | x_{i1})$. We will also use the letter q to denote conditional and unconditional probability distributions.

Independent pixels case

For Figure 1 the joint model for the observed and unobserved images can be written as

$$p(\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3 \mathbf{x}_4 \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z}) = \prod_{i=1}^n \prod_{t=1}^4 p(y_{it} | x_{it}) p(x_{it} | x_{it-1} z_i) p(z_i) \quad (1)$$

where we define $p(x_{i1} | x_{i0} z_i) = p(x_{i1} | z_i)$. To predict the true label, at each pixel i , we calculate $p(x_{it} | y_{i1} y_{i2} y_{i3} y_{i4} z_i)$ for $t=1,4$ using the efficient algorithm of Lauritzen and Spiegelhalter [11]. This produces a distribution for the true value at each pixel i . A class label can be selected, for example, by choosing the label with the maximum probability.

3.1 Handling spatial dependence

We construct a Gibbs distribution [12], [13] for all the images by combining the model in (1) above with a Markov random field model for each of the images. Let $\mathbf{x}_{r(i)s} = (x_{ks}, k=1, \dots, n \text{ and } k \neq i)$ i.e. the rest of the

labels excluding the label at pixel i in image s and define $\mathbf{x}_{n(i)s}$ to be the set of labels of the eight nearest neighbours of pixel i in image \mathbf{x}_s . For $s = 1, \dots, 4$ let $p(\mathbf{x}_s)$, $p(\mathbf{y}_s)$ and $p(\mathbf{z})$ be strictly positive distributions with the property that for each pixel i

$$p(x_{is} | \mathbf{x}_{r(i)s}) = p(x_{is} | \mathbf{x}_{n(i)s}) \quad (2)$$

and similarly for $p(\mathbf{y}_s)$ and $p(\mathbf{z})$. The notion here is that the label at a pixel interacts with its nearest neighbours, see [12] and [9]. Given the above, the joint model is defined as

$$q(\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3 \mathbf{x}_4 \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z}) = \exp\{ \log F + \log p(\mathbf{z}) + \sum_{s=1}^4 [\log p(\mathbf{x}_s) + \log p(\mathbf{y}_s)] \} / U \quad (3)$$

where F is the expression on the right hand side of Equation (1) and is a function of the images $(\mathbf{x}_s, \mathbf{y}_s, s=1,4)$ and \mathbf{z} . U is a normalising constant. The idea is to combine the CPN model of Figure 1 with models which encourage clustering of the map classes in each of the maps.

From (3), we compute $q(\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3 \mathbf{x}_4 | \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z})$ by the defining ratio and seek to find a set of class labels for the unobserved images with maximum probability given the observed images. Note that if the ordering is not important (with a slight abuse of notation) we can write

$$(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, \mathbf{x}_{r(i)1}, \mathbf{x}_{r(i)2}, \mathbf{x}_{r(i)3}, \mathbf{x}_{r(i)4}).$$

Hence by Bayes rule

$$\begin{aligned} q(\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3 \mathbf{x}_4 | \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z}) &= q(x_{i1} x_{i2} x_{i3} x_{i4} | \mathbf{x}_{r(i)1} \mathbf{x}_{r(i)2} \mathbf{x}_{r(i)3} \mathbf{x}_{r(i)4} \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z}) q(\mathbf{x}_{r(i)1} \mathbf{x}_{r(i)2} \mathbf{x}_{r(i)3} \mathbf{x}_{r(i)4} | \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z}) \\ &= q(x_{i1} x_{i2} x_{i3} x_{i4} | \mathbf{x}_{n(i)1} \mathbf{x}_{n(i)2} \mathbf{x}_{n(i)3} \mathbf{x}_{n(i)4} \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z}) q(\mathbf{x}_{r(i)1} \mathbf{x}_{r(i)2} \mathbf{x}_{r(i)3} \mathbf{x}_{r(i)4} | \mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \mathbf{y}_4 \mathbf{z}) \end{aligned} \quad (4)$$

The second line of equation (4) follows on computing the defining ratio from (3) and using the Markov property (2). Another way of thinking about this is that Equation (3) defines a vector valued Markov random field.

From (3) and (2), the first term on the right hand side of (4) above is proportional to

$$\prod_{t=1} p(y_{it} | x_{it}) p(x_{it} | x_{it-1} z_i) p(x_{it} | \mathbf{x}_{n(i)t}) \quad (5)$$

The factorisation in (5) corresponds to a CPN with graph as in Figure 5 below.

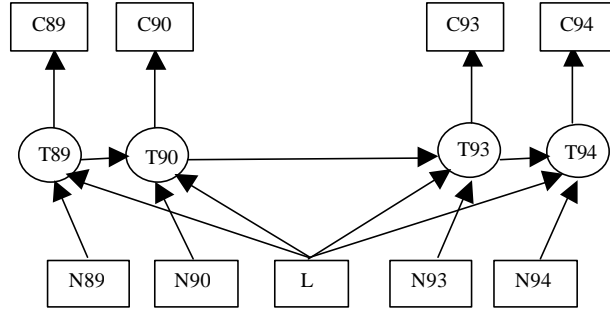


Figure 2: Modified conditional probability network which allows for spatial dependence

Note that this is simply Figure 1 augmented with some "dummy" neighbour nodes N89, N90, N93 and N94. It follows that existing algorithms for implementing the calculations of [11] in the case of independent pixels can be used here to calculate the appropriately normalised version of (5). All that is required is minor modifications to handle local neighbour information in the images, see also [2] and [6].

A cyclic ascent algorithm for doing the maximisation proceeds as follows:

1. start with initial estimates of the unobserved images $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ e.g. by using the results for the independent pixels case
2. visit each pixel in turn and at pixel i use (5) to compute

$$q(x_{i1} x_{i2} x_{i3} x_{i4} | \mathbf{x}_{n(i)1} \mathbf{x}_{n(i)2} \mathbf{x}_{n(i)3} \mathbf{x}_{n(i)4} y_{i1} y_{i2} y_{i3} y_{i4} z_i) \quad (6)$$

Choose map classes $x_{i1} x_{i2} x_{i3} x_{i4}$ to maximise (6). Another strategy at this point is to choose labels for the unobserved maps individually by maximising the **marginal** distributions

$$q(x_{it} | \mathbf{x}_{n(i)1} \mathbf{x}_{n(i)2} \mathbf{x}_{n(i)3} \mathbf{x}_{n(i)4} y_{i1} y_{i2} y_{i3} y_{i4} z_i)$$

for $t=1, \dots, 4$. The results presented here use this approach.

3. Continue cycling over all pixels until convergence.

To use the algorithm in practice requires the specification of parameters in the model. The specific Markov random field model used here was

$$p(x_{is} | \mathbf{x}_{n(i)s}) \propto \exp\{\alpha_s + \beta_s N(x_{is})\} \quad (7)$$

where $N(x_{is})$ is the number of 8 nearest neighbours of pixel i with label x_{is} , see for example [9]. For illustrative purposes we have used $\alpha_s = 0$ and $\beta_s = 1$ for all s . These parameters can be varied to change the relative weighting between image data and contextual information. The probabilities $p(y_{it} | x_{it})$ are assumed to be the same for each pixel and are estimated from error rates derived from the classification process. The remaining probabilities in (1) are estimated by the EM algorithm [2], [7] ignoring spatial dependence. Alternative estimation procedures such as coding or pseudo likelihood could also be used [9], [14].

The algorithm of [15] can be used for models with continuous-valued images and mixtures of discrete and continuous valued images with spatial dependence. We hope to report results at a later date.

4. Results

We used the CPN defined by Figure 1 to predict the true ground cover images (maps) T89, T90, T93 and T94, given the classified images C89, C90, C93, C94 and the landform L. Uncertainties in each of these input images were propagated through the CPN, giving posterior probability images for each of the “true”

maps. We also ran the CPN with and without spatial correlation included. For illustrative purposes, the most likely class maps for T93 are given in Figures 3 (without spatial correlation) and 4 (with spatial correlation) below, with roads overlaid in black. Note the reduction in “speckle” in Figure 4.



Figure 3: 1993 ground cover predicted by temporal CPN

The classification accuracies for an *independent* (i.e. not used in any model fitting) set of validation data are given in Table 1 below. This data consists of sites spread about the diagonal of the study area within an area occupying approximately one third of the region. In the table, spatial classification refers to the method defined in [9] and [10], but ignoring landform; the temporal model is the CPN of Figure 1 ignoring spatial correlation; and the spatial-temporal model is a CPN corresponding to Figure 1 which includes spatial correlation amongst pixels in the true images. Compared to the other models, the spatial-temporal model gives better classification accuracies.



Figure 4: 1993 ground cover predicted by spatial-temporal CPN

	Spatial classification		Temporal CPN		Spatial - Temporal CPN	
	salt	not salt	salt	not salt	salt	not salt
Aug 89	73.97	80.15	68.95	90.77	73.22	92.11
Sep 90	61.76	84.02	70.79	88.55	74.31	90.25
Sep 93	72.89	86.04	73.56	87.34	76.57	88.86
Aug 94	77.99	77.04	81.34	71.32	80.94	84.77

Table 1: Accuracy of classifications

The estimated percent area of salt derived from the spatial-temporal CPN model is 8.6% in 1989, 9.4% in 1990, 10.1% in 1993 and 12.2% in 1994.

A table of pixel counts for all possible combinations of sequences through time is given in Table 2.

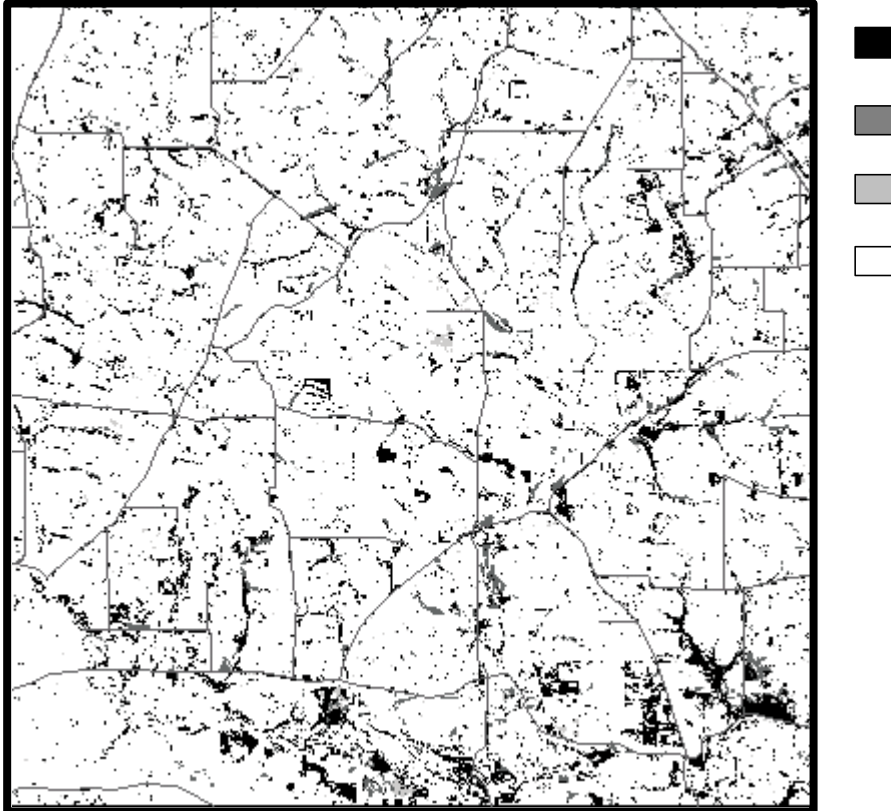
A few classes from the corresponding 16-class image are shown in Figure 4.

Aug 89	Sep90	Sep93	Aug94	# pixels	%	Comments
not salt	not salt	not salt	not salt	956354	87.576	(consistently not salt)
not salt	not salt	not salt	salt	22374	2.048	(turned salt in 94)
not salt	not salt	salt	not salt	48	0.004	
not salt	not salt	salt	salt	9670	0.885	(turned salt in 93)
not salt	salt	not salt	not salt	201	0.018	
not salt	salt	not salt	salt	31	0.002	
not salt	salt	salt	not salt	44	0.004	
not salt	salt	salt	salt	9547	0.874	(turned salt in 90)
salt	not salt	not salt	not salt	299	0.027	(turned not salt in 90)
salt	not salt	not salt	salt	219	0.020	
salt	not salt	salt	not salt	0	0.000	
salt	not salt	salt	salt	101	0.009	
salt	salt	not salt	not salt	2110	0.193	(turned not salt in 93)
salt	salt	not salt	salt	133	0.012	
salt	salt	salt	not salt	134	0.012	(turned not salt in 94)
salt	salt	salt	salt	90760	8.311	(consistently salt - black)

Table 2: Counts for all possible salt/not salt sequences derived from predicted maps

A priori, sequences with more than one salt/not salt transition are unlikely, since ground which is salt-affected is likely to stay that way. Table 2 shows that the predicted frequencies of such transitions are very small. Hence the model has succeeded in producing the sort of temporal consistency we would expect. Saline land which becomes non-salt-affected is rare, so that these pixels are likely to be incorrectly classified as salt initially; however a focus on genuine cases of these pixels may also provide insight into the land rehabilitation process.

Figure 5: map of all temporal sequences



5. Conclusions

Sensor derived data and expert temporal and contextual data can be combined in a rigorous framework using CPN's. The example considered suggests that CPNs are a useful tool for integrating data for environmental monitoring problems. They enable the effective combination of uncertain information from many different sources and model parameters can easily be estimated from the data and/or supplied by experts in the area. The calculations required by the models can be performed by an efficient algorithm [11]. Our experience has shown that analysing sequences of full TM scenes with the methods of this paper is computationally feasible.

6. References

1. Jensen, F. V. *An Introduction to Bayesian Networks*. Springer Verlag, New York , 1996.
2. Caccetta, P. *Remote Sensing, GIS and Bayesian Knowledge-based Methods for Monitoring Land Condition*. PhD thesis, Department of Computer Science, Curtin University of Technology, Western Australia, 1997.
3. Kiiveri, H. T. and Caccetta, P. Data integration, uncertainty and causal probabilistic networks for environmental monitoring. Submitted for publication.
4. Kiiveri, H. T. (1997) Assessing, representing and transmitting positional uncertainty in maps. *International Journal of Geographical Information Science*, 1997, 11, 1, pp. 33-52.
5. Badsberg, J. H. Model search in contingency table by CoCo. In Dodge, Y. and Whittaker, J., editors, *Computational Statistics, COMPSTAT 1992*, Neuchatel, Physica Verlag: Heidelberg, Vol. 1, 1992, pp. 251 -256.
6. Kiiveri, H. T. Some statistical models for remotely sensed data. In *SISC96 Imaging Interface Workshop Proceedings, 1996*.
7. Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39, 1977, pp.1-21.
8. Caccetta, P., Campbell, N., West, G., Kiiveri, H., and Gahegan, M. Aspects of reasoning with uncertainty in an agricultural GIS environment. *The New Review of Applied Expert Systems* 1, 1995, pp. 161-177.

9. Besag, J. E. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B* 48, 1986, pp. 259-302.
10. Kiiveri, H. T. and Campbell, N. A. Allocation of remotely sensed data using Markov models for image data and pixel labels. *Australian Journal of Statistics* 34, 1992, pp. 361-374.
11. Lauritzen, S. L., and Spiegelhalter, D. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society B* 50, 1988, pp. 157-224
12. Geman, S. and Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 1984, pp.721-741.
13. Darroch, J.N., Lauritzen, S.L. and Speed, T.P. Log-linear models for contingency tables and Markov fields over graphs. *Annals of Statistics* 8, 1980, pp. 522-539.
14. Besag, J. E. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, 36, 1974, pp. 192-326.
15. Lauritzen, S. L. Propagation of probabilities, means, and variances in mixed graphical association models. *Journal of the American Statistical Association* 87, 1992, pp. 1098-1108

About the Authors

Harri Kiiveri, BSc Ph.D obtained his Ph.D in Statistics from the University of Western Australia in 1983. He is a Principal Research Scientist with CSIRO Mathematical and Information Sciences. Recent work includes the development of methods for the assessment, representation and propagation of uncertainty in GIS. He is currently working on the theory and application of conditional probabilistic networks for image and non-image data. Harri is also involved with the practical application of methods for environmental monitoring, particularly for mapping and monitoring dryland salinity.

Peter Caccetta, BEng Ph.D received his Degree in Electrical Engineering from the University of Western Australia in 1989, and in 1997 his Ph.D in Computing Science from Curtin University of Technology, Perth, Western Australia. His activities since graduating in 1989 have been in the fields of remote sensing, geographical information systems and Bayesian knowledge based methods for monitoring land condition. His current interests include various aspects of image segmentation, environmental monitoring and machine learning.