An Object-oriented Spatial and Temporal Bayesian Network for Managing Willows in an American Heritage River Catchment

Lauchlin Wilkinson
Faculty of IT,
Monash Univ., AUS

Yung En Chee
School of Botany,
Univ. of Melbourne, AUS

Ann E. Nicholson
Faculty of IT,
Monash Univ., AUS

Pedro Quintana-Ascencio
Department of Biology,
Univ. of Central Florida, USA

Abstract

Willow encroachment into the naturally mixed landscape of vegetation types in the Upper St. Johns River Basin in Florida, USA, impacts upon biodiversity, aesthetic and recreational values. To control the extent of willows and their rate of expansion into other extant wetlands, spatial context is critical to decision making. Modelling the spread of willows requires spatially explicit data on occupancy, an understanding of seed production, dispersal and how the key life-history stages respond to environmental factors and management actions. Nicholson et al. (2012) outlined the architecture of a management tool to integrate GIS spatial data, an external seed dispersal model and a state-transition dynamic Bayesian network (ST-DBN) for modelling the influence of environmental and management factors on temporal changes in willow stages. That paper concentrated on the knowledge engineering and expert elicitation process for the construction and scenario-based evaluation of the prototype ST-DBN. This paper extends that work by using object-oriented techniques to generalise the knowledge organisational structure of the willow ST-DBN and to construct an object-oriented spatial Bayesian network (OOSBN) for modelling the neighbourhood spatial interactions that underlie seed dispersal processes. We present an updated architecture for the management tool together with algorithms for implementing the dispersal OOSBN and for combining all components into an integrated tool.

1 INTRODUCTION

The highly-valued Upper St. Johns River in Florida, USA has been the focus of considerable restoration investment (Quintana-Ascencio et al., 2013). However, woody shrubs, primarily Carolina willow (*Salix caroliniana* Michx.), have invaded areas that were historically herbaceous marsh (Kinser et al., 1997). This change to the historical composition of mixed vegetation types is considered undesirable, as extensive willow thickets detract from biodiversity, aesthetic and recreational values. Overabundance of willows reduces local vegetation heterogeneity and habitat diversity. People also prefer open wetlands that offer a viewshed, navigable access and scope for recreation activities such as wildlife viewing, fishing and hunting.

Managers seek to control the overall extent of willows, their rate of expansion into other extant wetland types and encroachment into recently restored floodplain habitats. Spatial context is critical to decision-making as areas differ in terms of biodiversity, aesthetic and recreational value, "invasibility" and applicable interventions. For instance, vegetation communities that are intact or distant from willow populations (seed sources) are less susceptible to invasion. With respect to interventions, mechanical clearing is restricted to areas where the substrate can support heavy machinery; prescribed fire depends on water levels and the quantity of "burnable" understorey vegetation.

Modelling willow spread requires spatially explicit data on willow occupancy, an understanding of seed production, dispersal, germination and survival, and how the key life-history stages respond to environmental factors and management actions. Data and knowledge on these pieces of the puzzle are available from ecological and physiological theory, surveys, field and laboratory experiments and domain experts.

State-transition (ST) models are a convenient means of organising information and synthesising knowledge to represent system states and transitions that are of

management interest. We build on recent studies that combine ST models with BNs to incorporate uncertainty in hypothesised states and transitions, and enable sensitivity, diagnostic and scenario analysis for decision support in ecosystem management (e.g. Bashari et al., 2009; Rumpff et al., 2011). Our approach uses the template described by Nicholson and Flores (2011) to explicitly model temporal changes in willow stages.

Nicholson et al. (2012) outlined the architecture of a management tool that would integrate GIS spatial data, a seed dispersal model and a state-transition dynamic Bayesian network (ST-DBN) for modelling the influence of environmental and management factors on temporal changes in willow stages. That paper described the knowledge engineering and expert elicitation process for the construction and scenario-based evaluation of the prototype ST-DBN. This paper extends that work, using object-oriented techniques to generalise the knowledge organisational structure of the willow ST-DBN and to construct an object-oriented spatial Bayesian network (OOSBN) for modelling the neighbourhood spatial interactions that underlie seed dispersal processes. We present an updated architecture for the management tool that incorporates GIS data and the new ST-OODBN and OOSBN structures, together with algorithms for implementing the dispersal OOSBN and for combining all components into an integrated tool.

2 BACKGROUND

Dynamic Bayesian Networks (DBNs) are a variant of ordinary BNs (Dean and Kanazawa, 1989; Nicholson, 1992) that allow explicit modelling of changes over time. A typical DBN has nodes for N variables of interest, with copies of each node for each *time slice*. Links in a DBN can be divided into those between nodes in the same time slice, and those in the next time slice. While DBNs have been used in some enviromental applications (e.g. Shihab, 2008), their uptake has been limited.

State-and-transition models (STMs) have been used to model changes over time in ecological systems that have clear transitions between distinct states (e.g., in rangelands and woodlands, see Bestelmeyer et al., 2003; Rumpff et al., 2011). Nicholson and Flores (2011) proposed a template for state-transition dynamic Bayesian networks (ST-DBNs) which formalised and extended Bashari et al.'s model, combining BNs with the qualitative STMs.

The influence of environmental and management factors on the main willow stages of management interest and their transitions is shown in our updated version of the Nicholson et al. (2012) ST-DBN (see Figure 1).

For each cell (spatial unit), data on attributes such as soil, vegetation type and information about landscape position and context is supplied from GIS data. This data provides inputs to parameterise the ST-DBN and dispersal model. A cell size of 100m x 100m (1 ha) was chosen to represent a modelling unit. This reflects the resolution of available spatial data for environmental attributes, makes the computational demand associated with seed dispersal modelling feasible, and is a reasonable scale with respect to candidate management actions. A time step of one year was considered appropriate given the willow's growth and seed production cycle (Nicholson et al., 2012).

Seed production depends on the size and number of reproductive (adult) stems *within* each cell. However, *Seed Availability*, the amount of seed available for germination within a cell, depends on willow seed production and dispersal from *surrounding* cells. As these processes are not accounted for in the ST-DBN (Figure 1), a key focus of this paper is the development and integration of an object-oriented spatial Bayesian network (OOSBN) to model the neighbourhood spatial interactions that underlie this process.

The purpose of the integrated tool is to synthesise current understanding and quantify important sources of uncertainty to support decisions on *where*, *when* and *how* to control willows most effectively. The ST-DBN models willow state transitions and characteristics in response to environmental and management factors within a single spatial unit and time step. For coherent, effective and well-coordinated landscape-scale management however, we want to be able to predict willow response across *space* (at every cell) in the target area and across *time* frames of management interest (e.g. 10-20 years). Such predictions can then be mapped and also aggregated across the target area to produce evaluation metrics for managers. Such a tool would enable managers to "test", visually compare and quantitatively evaluate different candidate management strategies.

This real-world management problem is naturally described in terms of hierarchies of components that include similar, repetitive structures. Object-oriented (OO) modelling has obvious advantages in this context. We apply OO techniques to generalise the knowledge organisational structure of the willow ST-DBN and design and construct the seed production and dispersal spatial network.

3 AN ST-OODBN FOR WILLOWS

Various authors have advocated the use of OO modelling techniques to: a) help manage BN complexity via abstraction and encapsulation, b) facilitate the
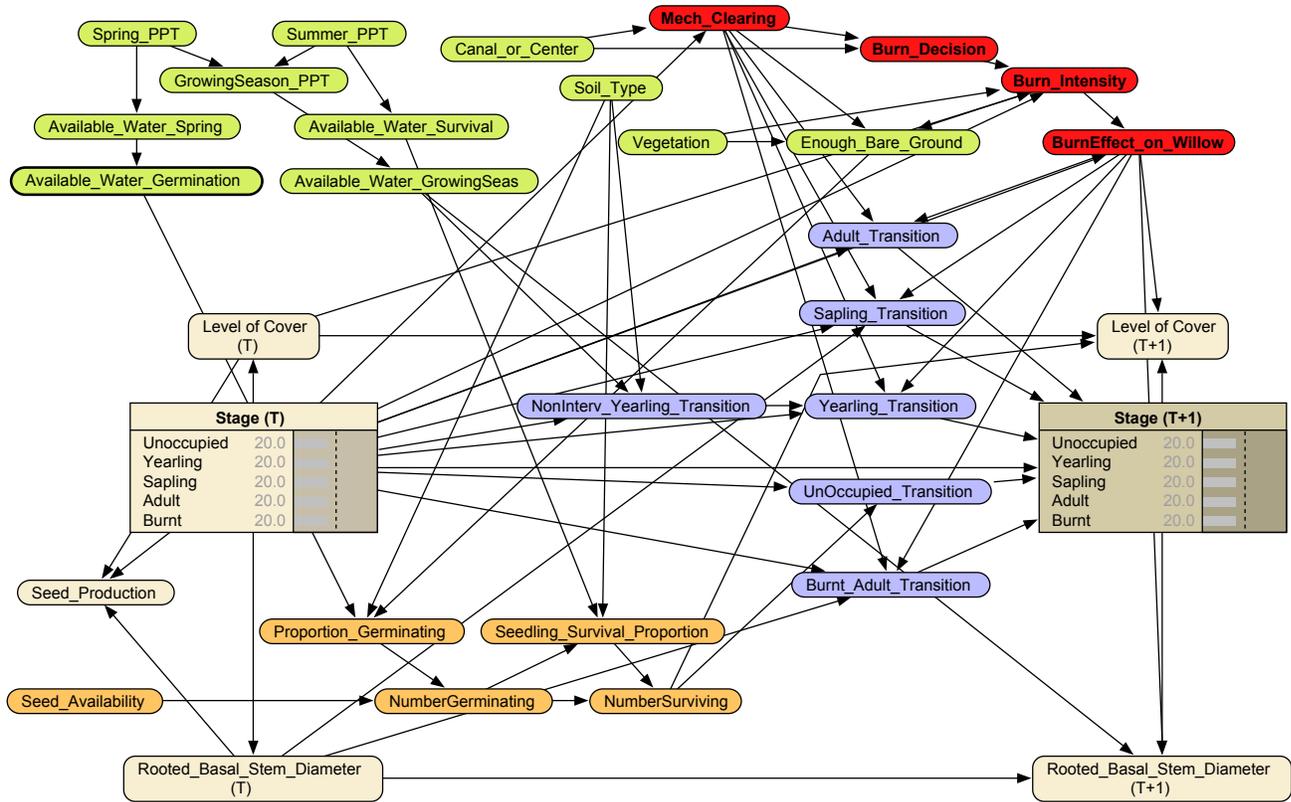
Figure 1: State-and-transition dynamic Bayesian network (ST-DBN) for modelling the response of key willow stages to environmental factors and management actions within a single spatial unit. The willow stages of management interest are: unoccupied, yearling, sapling, adult and burnt adult. Colours indicate: (i) aspects of willow state in tan; (ii) seed availability, germination and seedling survival processes in orange; (iii) environmental factors in green; (iv) management options in red; and (v) willow state-transitions in purple.

construction of classes of objects that are internally cohesive and potentially more reuseable, and c) formalise interfaces prior to integration (Koller and Pfeffer, 1997; Neil et al., 2000; Kjærulff and Madsen, 2008; Korb and Nicholson, 2010; Molina et al., 2010). However, examples in ecological and environmental management are scant (Molina et al., 2010; Carmona et al., 2011; Johnson and Mengersen, 2012).

We follow the definition of OOBNs used in Kjærulff and Madsen (2008), and implemented in the Hugin BN software package. A standard BN is made up of ordinary nodes, representing random variables. An OOBN class is made up of both nodes, and objects, which are instances of other classes. Thus an object may *encapsulate* multiple sub-networks, giving a composite and hierarchical structure. Objects are connected to other nodes via some of its own ordinary nodes, called its *interface* nodes. The rest of the nodes are not visible to the outside world, thus hiding information detail, another key OO concept. A class can be thought of as a self-contained 'template' for an OOBN object, described by its name, its interface and its hidden part.

Finally, interface nodes are divided into input nodes and output nodes. Input nodes are the root nodes within an OOBN class, and when an object (instance) of that class becomes part of another class, each input node may be mapped to a single node (with the same state space) in the encapsulating class. The output nodes are the only nodes that may become parents of nodes in the encapsulating class. When displaying an OOBN, we show Hugin[1] screen shots, where input nodes are indicated with a dotted and shaded outline, and output nodes with a bold and shaded outline.

We converted the ST-DBN (Figure 1) into a ST-OODBN as follows. Using the five conceptual categories of nodes from the original network as a guide, the network was split into two abstract class types. The first represents abstract influencing factors and consists of three sub-classes that define *Environmental Conditions*, *Management Options* and the germination and seedling survival *Processes Factors*. The second type represents state transitions of willows

---

[1]Note that Hugin OOBNs do not support inheritance, so there are no super-classes or sub-classes.

(see Figure 2) and defines five sub-classes, *TransitionFromUnoccupied*, *TransitionFromYearling*, *TransitionFromSapling*, *TransitionFromAdult* and *TransitionFromBurntAdult*. Figure 3 illustrates the definition of the *TransitionFromYearling* class.
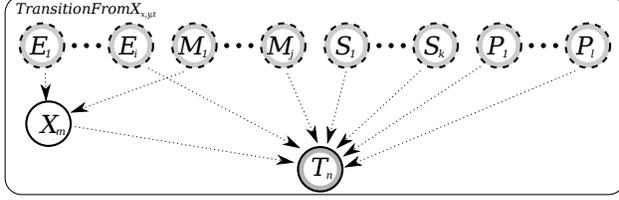

Figure 2: An abstract OOBN class for state transitions. Each implementation of a state transition output node T_n is defined by a combination of environmental conditions $E_{1...i}$, management options $M_{1...j}$, previous state variables $S_{1...k}$, process factors $P_{1...l}$ and any number of X_m hidden nodes. The only required input is the node that defines the previous state, all others are optional and are based on the implementing class. Dotted arrows indicate possible connections between input, hidden and output nodes.
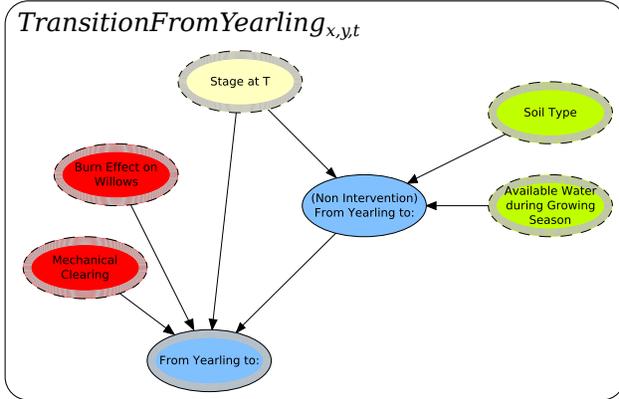

Figure 3: The $TransitionFromYearling$ implementation of the abstract $TransitionFromX$ type showing *Management Options* in red, *State Variables* in yellow and *Environmental Conditions* in green. The output *Transition* node *Transition from Yearling to:* is shown in blue with a bold and shaded outline.

These classes are instantiated as objects within a ST-OODBN class and when integrated with the seed dispersal OOSBN (described below) defines the complete ST model over a single time step (Figure 4). Recasting the network as a ST-OODBN makes the knowledge organisational structure explicit, whilst allowing network complexity to be hidden and integration efforts to focus on the interfaces between components of the network. Note that in the ST-OODBN (Figure 4)

Seed Availability is an input node. *Seed Production* and its spatial dispersal is modelled by a separate OO network, which we present next.

4 AN OOSBN FOR SEED PRODUCTION AND DISPERSAL

S.caroliana flowers in early spring and produces very large numbers of small seeds ($\sim$165,000 per average adult)that disperse by wind and water. Seed production is modelled by the *Willow Seed Production* OOBN (Figure 5), which is embedded in a broader seed dispersal model described below.

The number of seeds produced by an adult is given by the product of the number of *Inflorescences*, the number of *Fruits per inflorescence* and the number of *Seeds per fruit*. *Fruits per inflorescence* and *Seeds per fruit* are defined by distributions estimated from empirical data. The number of *Inflorescences* increases as a function of adult size (represented by *Rooted Basal Stem Diameter*) and this relationship has also been estimated from empirical data.

Cover is the percentage of a 1 hectare cell that is occupied by willows and *Average Canopy Area* is modelled as a function of *Rooted Basal Stem Diameter*. Together these two variables provide an estimate of the number of reproductive stems. Overall seed production within a cell, *Seeds per Hectare*, is then simply the product of the seed production per stem, by the number of reproductive stems. This *Willow Seed Production* OOBN models seed production processes explicitly rather than implicitly as in the ST-DBN prototype (Figure 1); an example of iterative and incremental knowledge engineering.

Willow seeds do not exhibit dormancy and have only a short period of viability – those that fail to germinate in the year they are produced are lost. The amount of seed available for germination within a cell depends on seed production and dispersal from surrounding cells. Thus, neighbourhood seed production and dispersal in combination with environmental and management factors determines patterns of willow spread and colonization.

Our approach to modelling seed dispersal is phenomenological rather than mechanistic. Wind-mediated seed dispersal is calculated using the Clark et al. (1999) dispersal kernel:

$$SD_{x,y}^{x',y'} = SP_{x',y'} \times \frac{1}{2\pi\alpha^2} e^{-(\frac{d}{\alpha})} \tag{1}$$

where $SD_{x,y}^{x',y'}$ is the number of seeds arriving at cell (x,y) from those produced at a cell (x',y'); it is the product of seed produced $SP_{x',y'}$ and an exponential
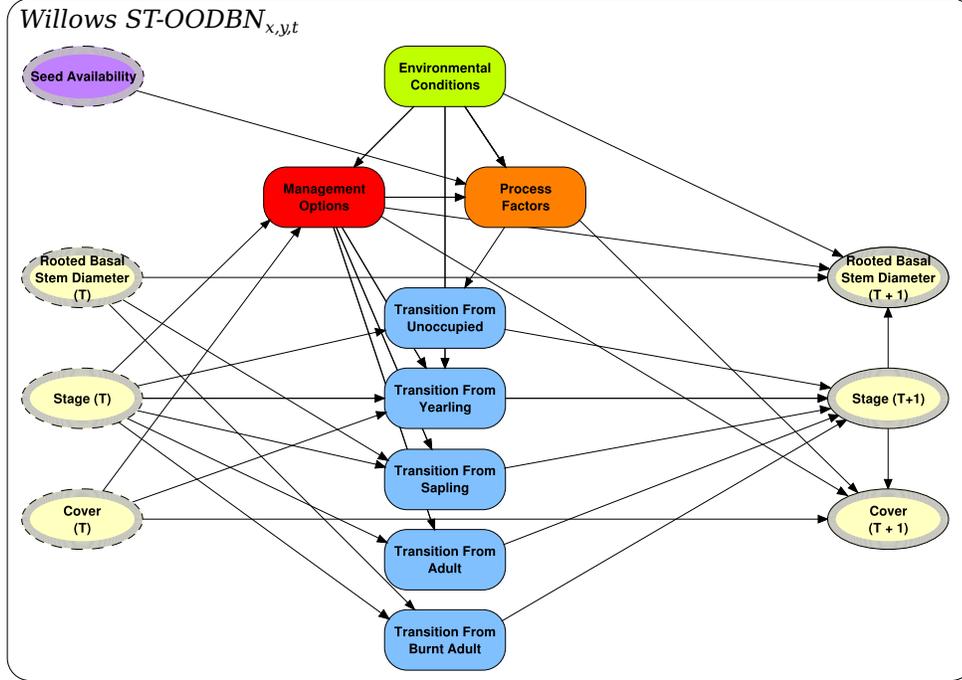
Figure 4: The resultant ST-OODBN class showing the four input nodes (*Seed Availability*, *Rooted Basal Stem Diameter (T)*, *Stage (T)* and *Cover (T)*) along with the *Environmental Conditions*, *Management Options*, *Process Factors* and five *TransitionFromX* objects that define the outputs (*Stage (T+1)*, *Rooted Basal Stem Diameter (T+1)* and *Cover (T+1)*) after one time step. Input nodes are illustrated with a dotted and shaded outline, instances of OOBN classes as round cornered boxes and output nodes with a bold and shaded outline.

kernel where d is the distance between cells (x, y) and (x', y'), and α is a distance parameter. To simulate stochasticity in dispersal events, α is a random variable that can be sampled from distributions designed to reflect the expected nature of dispersal (e.g. short versus long distance dispersal) (Fox et al., 2009). This seed dispersal model is captured within the *WindDispersalKernel*$_{x',y',t}$ OOBN (Figure 5), where the input *Distance* node is set for the particular (x, y) and α is set as a discretised normal distribution with a mean of 1 and a variance of 0.25.

For our purposes, we want to compute the seed availability $SA_{x,y}$ for a target cell (x, y), which is the sum of the seeds dispersed to it from every cell in the study area:

$$SA_{x,y} = \sum_{x',y' \in Area} SD_{x,y}^{x',y'} \qquad (2)$$

A naive BN model of this additive function would mean a *Seed Availability* node with all the *Seeds Dispersed* nodes (one for every cell) as its parents! For a study area with width w cells, height h cells, a *Seed Availability* node discretized to n states, and the *Seeds Dispersed* node discretized to m states, the CPT for *Seed Availability* would include $n \times m^{w \times h}$ probabilities –clearly infeasible.

From an ecological perspective, however, not *all* cells within a study area are expected to contribute towards final *Seed Availability* at any given cell. Indeed, because the number of seeds dispersed from a seed producing cell declines exponentially with increasing distance from that cell, we can make a simplifying assumption that after a certain distance, the number of seeds dispersed is effectively negligible. As a starting point we assume a circular region of influence for the target cell, defined by a dispersal mask with radius r. So for instance, a radius of eight cells (800 meters) implies $\pi 8^2$, or ~ 201 cells providing parents to the final *Seed Availability* node. This is still far too many, particularly as we are using a standard BN software package with discrete nodes and exact inference.

However, since *Seed Availability* is a simple additive function, we use the simple modelling trick of adding the *Seed Availability* from each cell to the cumulative seed availability so far (via node *Cumulative Seed Availability*). This is equivalent to repeatedly divorcing parents to reduce the size of the state space. We can think of this is as sequentially scanning over the spatial dimension in a similar way to rolling out a DBN over time. We call this a spatial Bayesian network (SBN), and the object-oriented variety an OOSBN.
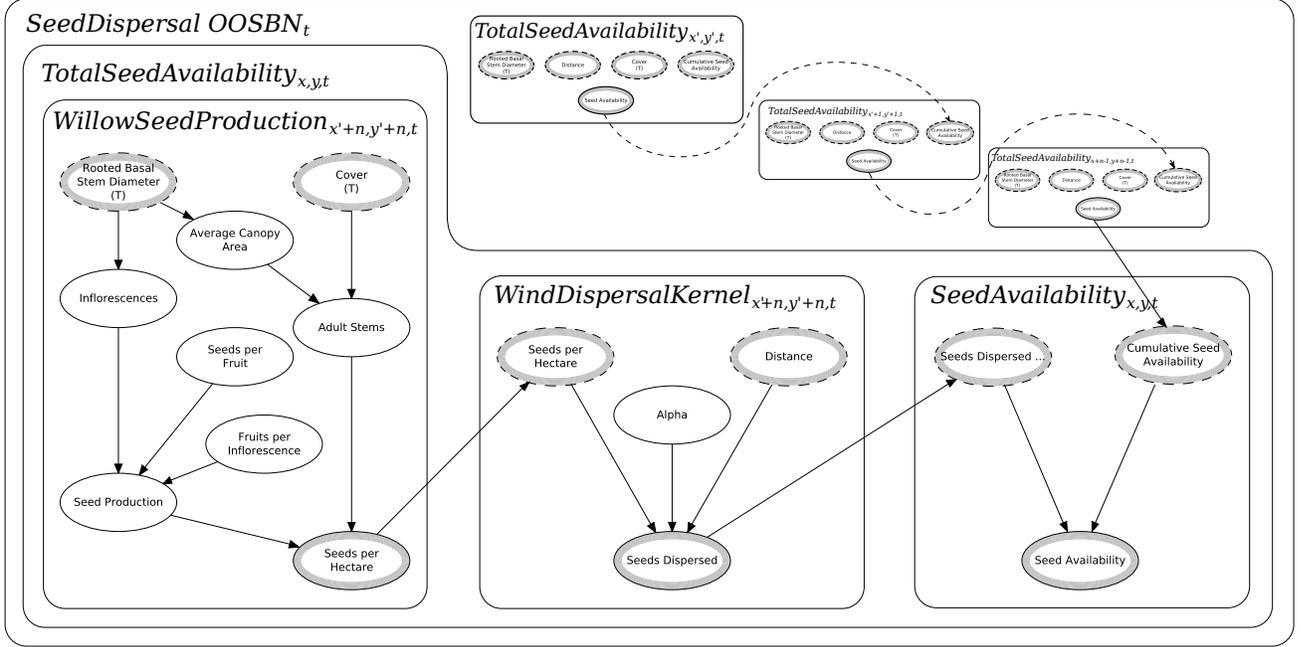
Figure 5: The OOSBN architecture showing how *Cover*, *Rooted Basal Stem Diameter*, *Cumulative Seed Availability* and *Distance* at locations (x', y') to $(x'+n, y'+n)$ are combined to provide total seed availability for each cell at location (x, y) at time t. Dashed arcs indicate that nodes are connected via the seed availability *PTLayer*. Multiple TotalSeedAvailability objects are illustrated within the SeedDispersal OOSBN reflecting that the total seeds available at a given (x, y) is dependant on seeds being produced in multiple locations. These synthetic TotalSeedAvailabilty objects are shown as collapsed objects, hiding their private nodes, and showing just the four input nodes that define the seeds dispersed from each producing cell to the target cell.

Our approach to integrating the seed dispersal makes use of such an OOSBN, that is run πr^2 times (i.e. once for every cell (x', y') within dispersal range) for each cell (x, y) in the study area to disperse and then sum the seed availability at each cell. The dispersal mask is flexible and can be designed to take on different shapes to reflect potentially important influences on wind dispersal such wind direction, wind strength and terrain characteristics. However, in the results given Section 6, we use a radius of eight cells.

Overall, our approach here is to mitigate the problem of large CPT sizes by turning the problem in to one of computation time. This has the added benefit of potentially being able to be computed easily in parallel and thus regaining some computation efficiency.

5 Integrating the ST-OODBN with the OOSBN

Figure 6 is an abstract representation of the system architecture, specifically, the interactions between the GIS layers and the OO networks as the tool is used for prediction at yearly intervals over the required management time frame.

For each cell in the GIS, there is conceptually one state-transition network (*ST-OODBN*) and one seed production and dispersal spatial network, (*OOSBN*). In practice, we do not store all these as separate networks, but rather re-use a single network structure, whose input nodes are re-parameterised for each cell, for each time step.

For the first time step, the system takes GIS data, which represents the initial conditions of the study area. These are stored in an internal data structure, *PTLayer*, that combines the spatial structure of the GIS, with distributions for the (discrete) nodes in the networks. *PTLayers* are used to store and pass the spatially referenced prior distributions of input nodes and posterior distributions of output nodes for both the *OOSBN* and *ST-OODBN* (Figure 6). Each *PTLayer* contains a number of fields, one for each of the node states of the linked input and output nodes. Each field stores the probability mass of the corresponding node state.

The *PTLayer* distributions are used to set the priors for the input nodes at each time step t. Then inference (belief updating) is performed within the OO network, producing new posterior probability distri-
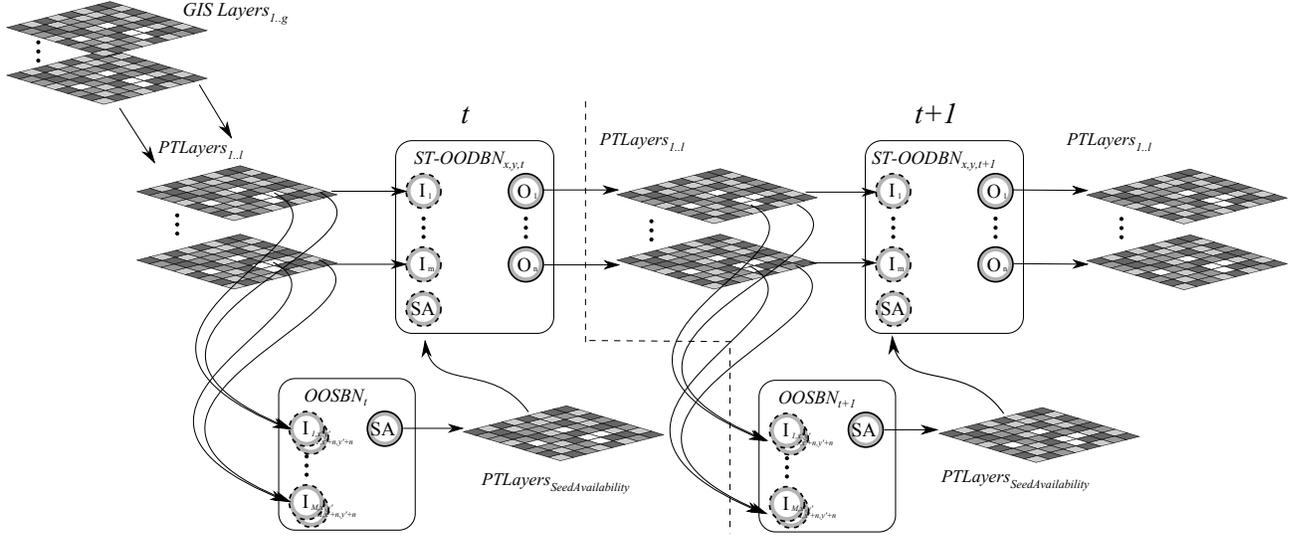
Figure 6: An abstract representation of the management tool architecture.

butions from the output nodes for each OO network, which are then transferred back to the corresponding *PTLayer*. This is equivalent to the standard "roll-out" followed by "roll-up" steps done in prediction with two time slice DBNs (Boyen and Koller, 1998) to avoid the computational complexity of rolling out a DBN over a large number of time steps. But here, in addition, the *PTLayers* are being used as an intermediate storage *across* the spatial grid, with the inputs for the next time $t + 1$ coming not only from both the network for the same cell, but from outputs from networks for *other* cells. This is done via the seed dispersal OOSBN, which uses the *Seed Availability PTLayer* to accumulate the seed availability arising from seed production in other cells within the dispersal mask. In effect, the *PTLayer* replaces both the temporal arcs if the network was rolled-out over many time steps, and the spatial arcs between the networks for different cells, which are essentially the cross-network arcs from seed production in one place to seed availability in another. Note that this method is limited to prediction only –we cannot use the model for diagnosis, or to identify the starting states and management actions to achieve a preferred end-state.

More formally, using the notation from Figure 6, at time t and at each cell location (x, y), $PTLayers_{1...l}$ are used to initialise the priors of nodes $I_{1...m}$ and $I_{1...M}$ of $ST\text{-}OODBN_{x,y,t}$ and $OOSBN_t$ respectively. After propagation within $OOSBN_t$, beliefs from the output node SA are stored in $PTLayer_{SeedAvailabilty}$ and then used to update the priors of input node SA of $ST\text{-}OODBN_{x,y,t}$. Then belief updating is done for the $ST\text{-}OODBN_{x,y,t}$ for each cell(x, y) and the beliefs from output nodes $O_{1...n}$ propagated back to $PTLay$-

$ers_{1..l}$ at time $t+1$. Although the mapping between the set of $PTLayers_{1...l}$ and input nodes $I_{1...m}$ is one-to-one, there may be cases where there are no GIS layers available for an input, in which case a prior distribution is used. Finally, with respect to OOSBN propagation, the range of locations $x', y' \ldots x' + n, y' + n$ (i.e. neighbourhood cells) that are included is defined by the dispersal mask described in the previous section.

Algorithm 1 An algorithm for propagating a GIS coupled ST-OODBN with spatial OOSBN sub-networks

---

1: **function** PROPAGATE($ST\text{--}OODBN, OOSBN, ptlayers, t$)
2:     $I \leftarrow I(ST\text{-}OODBN)$
3:     $O \leftarrow O(ST\text{-}OODBN)$
4:     $SA \leftarrow \text{getLayer}(ptlayers, \text{SeedAvailability})$
5:     **for** $t := 0$ **to** t **do**
6:         PROPAGATE(OOSBN,ptlayers,dispMask)
7:         // computes all Seed Availabilities
8:         **for all** $(x, y) \in Area$ **do**
9:             $p(SeedAvailability) \leftarrow SA_{x,y}$
10:             **for all** $I_i \in I$ **do**
11:                 $L_j \leftarrow getLayer(ptlayers, I_i)$
12:                 $p(I_i) \leftarrow L_j(x, y)$
13:             **end for**
14:             update beliefs in ST-OODBN
15:             **for all** $O_i \in O$ **do**
16:                 $L_j \leftarrow getLayer(ptlayers, O_i)$
17:                 $L_j(x, y) \leftarrow Bel(O_i)$
18:             **end for**
19:         **end for**
20:     **end for**
21: **end function**

The process is detailed in Algorithm 1, as a function *PROPAGATE* which takes the ST-OODBN (shown in Fig. 4), and OOSBN (shown in Fig. 5) networks, a list of *ptlayers* (previously initialised from the GIS layers) whose cells correspond to the area under consideration, and the number of time steps T over which to propagate the network. In the algorithm, we use

$I(OOBN)$ (respectively $O(OOBN)$) to denote a function that returns the input (resp. output) nodes of the interface of the OOBN, and $getLayer(ptlayers, V)$ to denote a function that returns the $PTLayer$ corresponding to a node V.

Algorithm 2 An algorithm for dispersing seeds by wind using an OOSBN class

---
```
1: function PROPAGATE(OOSBN, ptlayers, dispMask)
2:     SA ← getLayer(ptlayers, SeedAvailability)
3:     for all (x, y) ∈ Area do
4:         P(SA_{x,y} = none) ← 1
5:     end for
6:     for all (x, y) ∈ Area do
7:         for all I_i ∈ I_{x,y}(OOSBN) do
8:             L_j ← getLayer(ptlayers, I_i)
9:             p(I_i) ← L_j(x, y)
10:        end for
11:        for all (x', y') ∈ dispMask do
12:            p(CumSeedAvail) ← SA_{x,y}
13:            d ← √((x − x')² + (y − y')²)
14:            p(Distance = d) ← 1
15:            update beliefs in OOSBN
16:            SA_{x,y} ← Bel(SeedAvailability)
17:        end for
18:    end for
19: end function
```

First, seed dispersal is done with the OOSBN using Algorithm 2. Then for each cell (x, y) in the area, for each input node I_i, the distribution for that cell from its corresponding $ptlayer$ is set as the prior of the input node. Belief updating of the ST-OODBN is done, propagating the new priors through to updated posterior distributions for the output nodes. Finally the beliefs for each output node ($Bel(O_i)$) are copied back to the distribution at the (x, y) cell for the corresponding $ptlayer$ (i.e. into $SA_{x,y}$).

Algorithm 2 details the propagation process using the *SeedDispersal OOSBN* class illustrated in Figure 5. The algorithm starts by taking the OOSBN, a list of $PTLayers$ whose cells correspond to the area under consideration, and a dispersal mask ($dispMask$). Before starting the dispersal process, the provided *Seed Availability PTLayer* is initialised with no seeds available at all (x, y) co-ordinates. It then loops through each cell (x, y) in the area, setting the distribution for each input node I_i (i.e., *Cover*, *Rooted Basal Stem Diameter* and *Cumulative Seed Availability*) from its corresponding $PTlayer$ cell. The algorithm then enters a second loop for every cell that is a possible seed source based on the dispersal mask. The *Distance* node is set using the Euclidean distance between the current cell and the target co-ordinates. Finally, belief updating is done within the OOSBN (Figure 5) and the beliefs (i.e. the posterior probability distribution) from the *Seed Availability* node at each point are transferred into the *Cumulative Seed Availability* for the subsequent cell at each iteration. After all the cells (x', y') within the dispersal mask have been visited, the *Seed Availability PTLayer*, $SA_{x,y}$ contains the overall seed availability in that cell, which is used for modelling germination in the ST-OODBN in Algorithm 1.

6 PRELIMINARY RESULTS

To implement the software architecture described we chose to use Hugin Researcher 7.7(2013) to develop the ST-OODBN and dispersal model OOSBN, the Hugin Researcher Java API 7.7 (2013) to provide programmatic access to the developed networks, the Image-IO-ext (2013) java library to provide access to GIS raster layer formats, and the Java programming language to implement the algorithms tying the components together. Hugin was chosen as the OOBN development platform as it currently has one of the most complete OOBN implementations. Java was chosen as the implementing language as it is platform independent and provides for a well established and understood OO development environment. We implemented the tool as a standalone program allowing preprocessing of GIS data to be performed in whatever program the end user was most familiar with. In our case we used a combination of ArcGIS (2013), Quantum GIS (2013)and SAGA GIS(2013).

To demonstrate our working implementation, we ran the model for the Blue Cypress Marsh Conservation Area (138 x 205 cells) within the Upper St. Johns River basin. We used a simplified (and unrealistic) management rule set that says if a cell is next to a canal, mechanical clearing is carried out, otherwise for landlocked cells, burning is prescribed (with a probability of 0.1). Maps of willow cover and seed production were generated at yearly intervals for a 25 year prediction window. This took about 8 hours of computation on a 64bit machine with a 2.8GHz processor. Figure 7 shows seed availability across the study area using output from the *SeedAvailability* nodes in the OOSBN at 5 yearly intervals. To produce the maps, the seed availability interval with the highest posterior probability distribution is used to produce a grayscale value where *zero* seeds is black, and 10^{12} is white. In the run shown, seed availability decreases over time as the level of willow cover is reduced by the management regime.

7 DISCUSSION AND FUTURE WORK

For coherent, coordinated and effective landscape-scale decision support, managers need the capability to predict willow state changes across *space* and *time*. We have tackled the challenges of this real-world problem by synthesising ideas and techniques from object-oriented knowledge engineering, dynamic BNs, GIS-
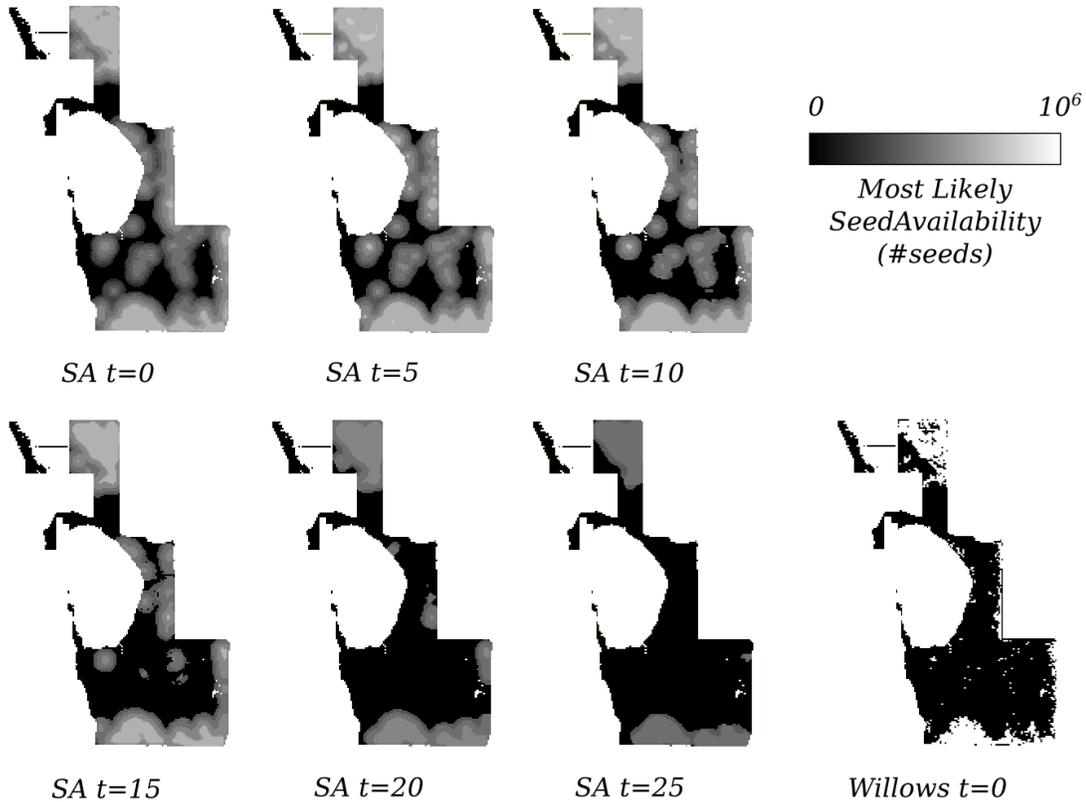
Figure 7: Seed availability predicted by the Willows ST-OODBN at $t = 0, 5, 10, 15, 20, 25$ years, across the Blue Cypress Marsh Conservation Area (138 x 208 cells). Adult willow occupancy at $t = 0$ is shown in the bottom right panel; black indicates absence, grey presence.

coupled BNs and dispersal modelling. To our knowledge, this is the first environmental management application in which OOBNs are used to model spatially-explicit process interactions.

Further work in the development of this management tool includes developing a water dispersal model and updating the parametrisation of the *ST-OODBN* and *OOSBN* using judgements from a larger pool of domain experts, together with specific empirical data where available. In addition, we will work with managers and domain experts to identify: i) realistic management scenarios, ii) useful summary descriptors for the various model outputs and iii) desirable features for a tool interface.

Throughout the research and integration process we encountered challenges with the development, management and use of OOBNs. While there have been advances in OOBN software, they still lack a lot of the useful features available in other development tools. For instance, modern software engineering IDEs provide easy to use re-factoring, documentation and integration with version control tools. The tools we used to design and implement the underlying OOBNs for

our tool still lack powerful refactoring, making the management of object interface changes a time consuming and error prone task. Integrated source control is non-existent and documentation tools rudimentary. Improvements in these areas would make working with OOBNs far more accessible to the type of user that wishes to make use of OOBNs for natural resource management.

With respect to spatialising the ST-OOBN with the use of OOSBNs, there is currently no graphical tool up to the task of facilitating the integration of the required components. This means that anyone wanting to replicate our work would need to make use of the available APIs and this constitutes a barrier to usage by people with no or little programming background.

Acknowledgements

References

Bashari, H., Smith, C., and Bosch, O. (2009). Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks. *Agricultural Systems*, 99(1):23–34.

Bestelmeyer, B. T., Brown, J. R., Havstad, K. M., Alexander, R., Chavez, G., and Herrick, J. E. (2003). Development and use of state-and-transition models for rangelands. *Journal of Range Management*, (2):114–126.

Boyen, X. and Koller, D. (1998). Tractable inference for complex stochastic processes. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, pages 33–42, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Carmona, G., Molina, J., and Bromley, J. (2011). Object-Oriented Bayesian networks for participatory water management: two case studies in Spain. *Journal of Water Resources Planning and Management*, 137(4):366–376.

Clark, J. D., Silman, M., Kern, R., Macklin, E., and Hille Ris Lambers, J. (1999). Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, 80(5):1475–1494.

Dean, T. and Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150.

Esri (2013). ArcGIS (Version 10.1) [Software]. `http://www.esri.com/software/arcgis`.

Fox, J. C., Buckley, Y. M., Panetta, F. D., Bourgoin, J., and Pullar, D. (2009). Surveillance protocols for management of invasive plants: modelling Chilean needle grass (*Nassella neesiana*) in Australia. *Diversity and Distributions*, 15(4):577–589.

GeoSolutions (2013). Image-IO-ext Java Library (Version 1.1.7) [Software]. `http://github.com/geosolutions-it/`.

Hugin Expert A/S (2013). Hugin Researcher (Version 7.7) [Software]. `http://www.hugin.com`.

Johnson, S. and Mengersen, K. (2012). Integrated Bayesian network framework for modeling complex ecological issues. *Integrated Environmental Assessment and Management*, 8(3):480–90.

Kinser, P., Lee, M. A., Dambek, G., Williams, M., Ponzio, K., and Adamus, C. (1997). Expansion of Willow in the Blue Cypress Marsh Conservation Area, Upper St. Johns River Basin. Technical report, St. Johns River Water Management District, Palatka, Florida.

Kjærulff, U. B. and Madsen, A. (2008). *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Springer Verlag, New York.

Koller, D. and Pfeffer, A. (1997). Object-oriented Bayesian networks. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, UAI'97, pages 302–313, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Korb, K. B. and Nicholson, A. (2010). *Bayesian artificial intelligence*. Chapman&Hall/CRC, Boca Raton, FL, 2nd edition.

Molina, J., Bromley, J., García-Aróstegui, J., Sullivan, C., and Benavente, J. (2010). Integrated water resources management of overexploited hydrogeological systems using Object-Oriented Bayesian Networks. *Environmental Modelling & Software*, 25(4):383–397.

Neil, M., Fenton, N., and Nielson, L. (2000). Building large-scale Bayesian networks. *The Knowledge Engineering Review*, 15(3):1–33.

Nicholson, A., Chee, Y., and Quintana-Ascencio, P. (2012). A state-transition DBN for management of willows in an american heritage river catchment. In Nicholson, A., Agosta, J. M., and Flores, J., editors, *Ninth Bayesian Modeling Applications Workshop at the Conference of Uncertainty in Artificial Intelligence*, Catalina Island, CA, USA, http://ceur-ws.org/Vol-962/paper07.pdf.

Nicholson, A. E. (1992). *Monitoring Discrete Environments using Dynamic Belief Networks*. PhD thesis, Department of Engineering Sciences, Oxford.

Nicholson, A. E. and Flores, M. J. (2011). Combining state and transition models with dynamic Bayesian networks. *Ecological Modelling*, 222(3):555–566.

Quantum GIS Development Team (2013). Quantum GIS (Version 1.8.0) [Software]. `http://www.qgis.org/`.

Quintana-Ascencio, P. F., Fauth, J. E., Castro Morales, L. M., Ponzio, K. J., Hall, D., and Snyder, K. (2013). Taming the beast: managing hydrology to control Carolina Willow (*Salix caroliniana*) seedlings and cuttings. *Restoration Ecology*.

Rumpff, L., Duncan, D., Vesk, P., Keith, D., and Wintle, B. (2011). State-and-transition modelling for adaptive management of native woodlands. *Biological Conservation*, 144(4):1224–1236.

SAGA Development Team (2013). SAGA GIS (Version 2.0.8) [Software]. `http://www.saga-gis.org/`.

Shihab, K. (2008). Dynamic modeling of groundwater pollutants with Bayesian networks. *Applied Artificial Intelligence*, 22(4):352–376.