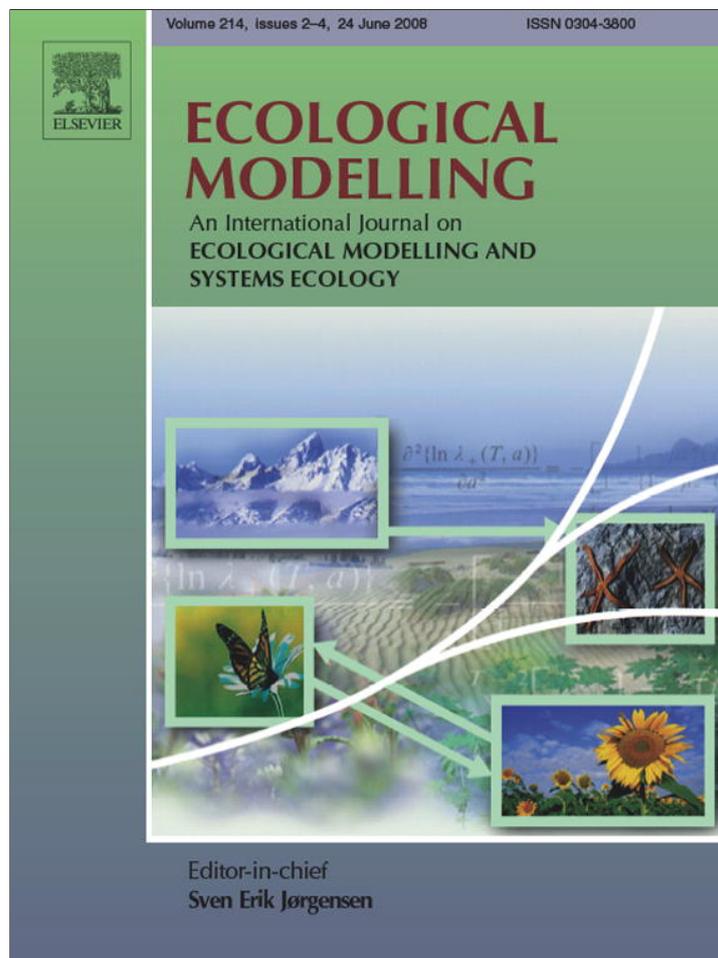


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

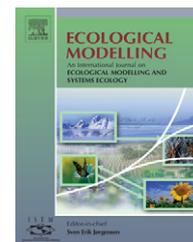
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

available at [www.sciencedirect.com](http://www.sciencedirect.com)journal homepage: [www.elsevier.com/locate/ecolmodel](http://www.elsevier.com/locate/ecolmodel)

# Monitoring amphibian populations with incomplete survey information using a Bayesian probabilistic model

Duncan S. Wilson<sup>a,\*</sup>, Margo A. Stoddard<sup>b</sup>, Klaus J. Puettmann<sup>a</sup>

<sup>a</sup> Department of Forest Science, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331, USA

<sup>b</sup> Department of Wildlife Ecology and Conservation, University of Florida, 110 Newins-Ziegler Hall, Gainesville, FL 32611-0430, USA

## ARTICLE INFO

### Article history:

Received 7 November 2007

Received in revised form

11 February 2008

Accepted 19 February 2008

Published on line 1 May 2008

### Keywords:

Bayesian networks

Hierarchical Bayesian models

Amphibian monitoring

*Dicamptodon tenebrosus*

*Rhyacotriton* spp.

*Ascaphus truei*

## ABSTRACT

Bayesian networks (BNs) are a probabilistic modeling platform that connect variables through a series of conditional dependences. We demonstrate their utility for broad-scale conservation of amphibian populations where different types of information may be available within the region. Wildlife conservation decisions for most species are made jointly with other objectives and are tightly constrained by finances. Bayesian networks allow the use of all available information in predictions, and can provide managers with the best available information for making decisions. Habitat models were developed as a hierarchical Bayesian (HB) model for aquatic amphibian populations in the temperate Oregon Coast Range, USA. Predictions for new streams sections were made jointly using a Bayesian network to allow the inclusion of different types of available information. Missing habitat variables were modeled based on habitat survey information. Uncertainty in the true (but unknown) habitat variables were incorporated into the prediction intervals. Further, the probabilistic approach allowed us to incorporate survey information for co-occurring species to help make better predictions. Such species information was connected through the Bayesian network by the conditional dependence that arises from shared habitat variables. The utility of Bayesian networks was shown for these populations for broad-scale risk management. In contrast to deterministic models, the probabilistic nature of Bayesian networks is a natural platform for incorporating uncertainty in predictions and inference.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

Wildlife management in the forested Pacific Northwest (PNW) is integrated with other objectives such as timber production and recreation. As such, most wildlife populations are considered with an informal risk management or adaptive management approach (FEMAT, 1993). With a risk management approach, decisions are partly based on perceived risk to a species which may be lower for certain types of management, or where the species is relatively abundant in adjacent areas.

Adaptive management uses population information and trends following large-scale habitat manipulations to adjust ongoing management strategies. These population trends cannot be determined, and decisions on whether mitigation is necessary cannot be made without broad-scale survey information for multiple species (Marsh and Trenham, 2001; Muths et al., 2006). Both management approaches therefore rely heavily on population information gathered through surveys.

The extent to which a wildlife population is monitored depends primarily on the conservation status of the species.

\* Corresponding author. Tel.: +1 541 737 6215; fax: +1 541 737 1393.

E-mail addresses: [duncan.wilson@oregonstate.edu](mailto:duncan.wilson@oregonstate.edu), [dunc.steil@yahoo.com](mailto:dunc.steil@yahoo.com) (D.S. Wilson), [mstodd@ufl.edu](mailto:mstodd@ufl.edu) (M.A. Stoddard), [klaus.puettmann@oregonstate.edu](mailto:klaus.puettmann@oregonstate.edu) (K.J. Puettmann).

0304-3800/\$ – see front matter © 2008 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2008.02.003

Federally listed, Endangered Species have formal monitoring programs in place, but Species of Concern for which more information are needed to determine their conservation status are monitored in a piecemeal fashion (e.g., FEMAT, 1993). An example is the Federal Survey and Manage program in the Pacific Northwest (Olson, 1999; Molina et al., 2006). Areas considered for management are intensively surveyed for plant and wildlife Species of Concern, and when found, appropriate mitigation is considered in management plans. Relatively abundant species are not considered, even though they may be rare or absent in a particular watershed or near the edge of their range.

Information gathered in the Survey and Manage program is area-specific and not formally used in management decisions made at the regional scale or even on adjacent land. This unfortunate shortfall is due to its *ad hoc* nature and is unavoidable with scarce monitoring funds that cannot cover all wildlife species or allow systematic continuous monitoring. Even so, such information has potential use and may be incorporated into risk and adaptive management strategies using probabilistic models (Varis and Kuikka, 1999; Borsuk et al., 2004; Lee and Irwin, 2005).

Our objective was to demonstrate the utility of a joint Bayesian parameter estimation and prediction approach for monitoring and risk management. The approach makes use of the probabilistic modeling framework of Bayesian networks (BNs). The manner in which BNs portray conditional relationships allows different types of evidence to be used in predictions. Missing covariates are modeled using a probability distribution of likely values. Further, predictions can incorporate survey information for co-occurring species when species are conditional dependent on a subset of the same habitat variables. This unique property of BNs allows 'information passing' between co-occurring species through the shared habitat variables. We further discuss how the BN probabilistic framework can be exploited by a diverse suite of modeling topics.

We illustrate the BN approach with data from a multi-scale amphibian study set in the Oregon Coast Range (Stoddard and Hayes, 2005). Amphibians are closely monitored in the PNW since some species are sensitive to habitat changes caused by logging (Corn and Bury, 1989; Vesely and McComb, 2002). In this study, habitat models for four aquatic species were developed as a hierarchical Bayesian (HB) model. In addition to new predictions based on model parameters and measured covariates, the BN framework allows predictions with incomplete information. We explore the utility of the approach for predicting probability of occurrence under different types of incomplete information, including missing habitat covariates. We also examine possible gains in survey efficiency in making predictions using survey information for co-occurring species. Finally, we discuss the benefits of a probabilistic model-based framework for risk management of these amphibian populations.

## 2. Study area and amphibian surveys

The study was conducted in drainages on Bureau of Land Management (BLM) land on the eastern side of the Oregon Coast

Range, USA. The region has a moist maritime climate, characterized by cool wet winters and dry summers. Forest stands within study drainages comprised primarily second-growth *Pseudotsuga menziesii* forests naturally regenerated following logging, or younger planted clearcuts. The study area lies within the low elevation, *Tsuga heterophylla* zone (Franklin and Dyrness, 1973).

Sixteen third-order drainages ranging from 90 to 200 ha were randomly sampled from the population of all drainages on BLM lands of the Eugene and Salem Districts. Amphibians were censused in 35–50 randomly chosen 2-m lengths of stream (sample points) within each drainage. Sample points had visible surface flow, were less than 0.5 m deep, and were located in ephemeral, discontinuous or perennial streams. Presence or absence was recorded of Pacific giant salamanders (*Dicamptodon tenebrosus*) (PGS), larval (LTF) and adult (ATF) tailed frogs (*Ascaphus truei*), southern torrent salamanders (*Rhyacotriton variegatus*) (TS), and Columbia torrent salamanders (*Rhyacotriton kezeri*) (TS). The two *Rhyacotriton* spp. were geographically separate and had similar habitat associations (Wilkins and Peterson, 2000; Stoddard and Hayes, 2005) and thus were combined for analysis. At least one species was found at 59% of the stream sections sampled. Methods for surveying amphibians followed standard protocols (Bury and Corn, 1991).

Sample points were distributed randomly within the drainages, but were later grouped into stands based on the age classes of forest bordering the stream (classes: 0–15, 16–55, 55–105 and greater than 105 years). Stands contained between 1 and 34 sample points, with a mean of 4.6. A total of 702 sample points across 153 stands were surveyed in the summers of 1998 and 1999. At each sample point, potential habitat variables were recorded including the percentage slow water (%SLOW) (pools and glide), percentage fine substrate (%FINE) (bedrock and particle sizes less than 32 mm), mean stream width (m) (WIDTH), percentage of undercut bank (%UNDER), and percentage cover of coarse wood (%CWD). Habitat variables at the stand- and drainage-scales were not considered explicitly in this analysis. A complete description of the study area and sampling methods is found in Stoddard and Hayes (2005).

## 3. Methods

### 3.1. Bayesian parameter estimation

Presence absence data were analyzed using an HB model (Gelman et al., 2004). At the stream-scale (2-m sample points) we used logistic regression to characterize a habitat model as

$$Y_{ijkl} \sim \text{Bern}(p_{ijkl}) \quad (1)$$

$Y_{ijkl}$  was the presence/absence data at sample point  $l$ , for species  $i$  in stand  $j$ , and drainage  $k$ . The probability of occurrence ( $p$ ) was modeled using logistic regression as

$$\text{logit}(p_{ijkl}) = \beta_{0i} + \alpha_{ijk} + \beta_{1i}(\%FINE) + \beta_{2i}(\%SLOW) + \beta_{3i}(\%WIDTH) + \beta_{4i}(\%UNDER) + \beta_{5i}(\%CWD)$$

All habitat variables were standardized prior to analysis by subtracting the mean and dividing by the standard deviation (S.D.). Non-informative priors were specified for each lower-level parameter ( $\beta_0 - \beta_5$ ) as  $\sim N(0, 100)$ . Initial covariate selection was guided by results from previous analyses of the data (Stoddard and Hayes, 2005). The influences of stand- and drainage-scale variables were included as higher-level covariates. Stand-scale effects were assumed to modify the intercept ( $\beta_0$ ) as

$$\alpha_{ijk} \sim N(\gamma_{ik}, \sigma_{\alpha_i}^2)$$

These predictions were modified at the drainage-scale in a hierarchical manner as

$$\gamma_{ik} \sim N(0, \sigma_{\gamma}^2).$$

A hierarchical structure was not imposed on the slope parameters ( $\beta_{1-5}$ ). All analyses were done with WinBUGS ver. 1.4.1 which uses a Markov chain Monte Carlo (MCMC) approach to characterize the posterior distributions (Lunn et al., 2000). Partially informative priors were used for the higher-level variance terms as  $\sigma_{\alpha}$  and  $\sigma_{\gamma} \sim \text{Unif}(0, 20)$ . This was done to avoid situations where the MCMC chain became ‘stuck’ at infinite variance estimates (Gelman and Hill, 2007). This was a reasonable assumption, since a one standard deviation range of 0–20 on the logit scale spans the range of probabilities found in this study. Specifying purely non-informative priors had no appreciable effect on the results (except for the problem mentioned). A common variance across species for the drainage effects was found to be an adequate assumption, judged by a deviance information criterion (DIC). For stand effects, PGS, JTF, and ATF were found to have a similar variance and were combined. TS showed a higher variance and this variance was estimated separately. Habitat covariate selection was also guided by a lowest DIC criterion in order to select a parsimonious model that still had high predictive ability. The final model with combined variances as indicated had the lowest DIC of alternative options. This study was concerned with stream-scale habitat models, so the higher-scale stand and drainage effects were considered nuisance variables. Results were based on 100,000 MCMC simulations with every 10th sample retained to reduce autocorrelation, following a 20,000 iteration burn-in.

### 3.2. Bayesian network predictions

The focus of this study was predicting amphibian occurrence for new sites when only partial habitat information was available. These predictions were made simultaneously with the model fitting within WinBUGS using the MCMC results for the model parameters. These simulated predictions were made for new stands in a new drainage. The stand and drainage effects were set to zero, which was the mean effect across all drainages. Predictions were made under different types and levels of missing information, with results grouped into three sections. First, we examined model predictions based on fully measured covariates, and these were compared to predictions made where certain covariates were missing. The next section examined the use of co-occurring species survey results for prediction with missing covariates. The last section examined combinations of co-occurring species and selected covariates that were measured with others missing.

Missing covariates were modeled with parametric distributions to approximate the shape and range of the data (shown in Fig. 1). It is important to distinguish that these covariates were missing by design, i.e., not part of the measurement protocol. Hence, each covariate was either completely measured at all sample points, or completely missing. The variables %FINE, %SLOW and %UNDER were modeled as being uniformly distributed across the range of observed values. This was a realistic assumption for the covariates, with the exception that %SLOW and %UNDER had slightly higher probability mass at the extremes (i.e., at 0 and 100%). An alternative approach would be to model covariates as categorical variables, with frequency of classes based on the observed data (Borsuk et al., 2004). Observations of %CWD and WIDTH showed a distinct mode, and were modeled (on the standardized scale) with a normal and exponential distribution, respectively. WIDTH was modeled  $\sim N(0,1)$ , and %CWD as  $\sim \text{Exp}(1.8)$ . Habitat covariates were not correlated, with the highest  $r=0.23$ .

New predictions were made with a Bayesian network (Pearl, 1988; Lee and Rieman, 1997; Reckhow, 1999; Borsuk et al., 2004) simultaneous to the Bayesian parameter estimation. BNs are graphical models used primarily to characterize systems probabilistically, and to update probabilities with new evidence (Pearl, 1988; Jensen and Nielsen, 2007). The graphical model

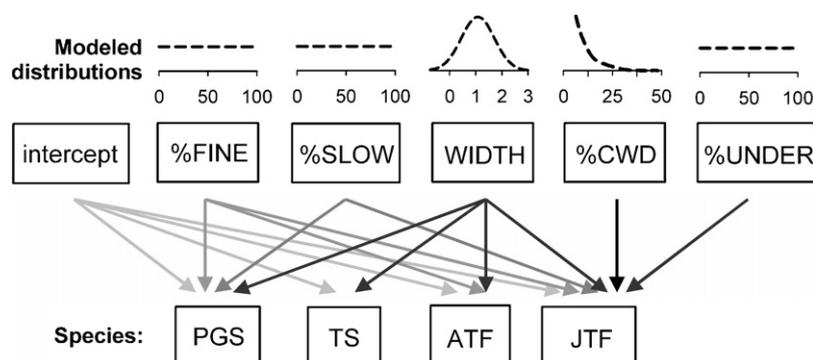


Fig. 1 – Bayesian network diagram with model structure indicated for variables (boxes) and conditional relationships (arrows). The intercept is further modified by stand- and drainage-scale effects (not shown). Habitat covariates were modeled (if missing) with parametric distributions shown at the top.

**Table 1 – Parameter estimates for Eq. (1), with standard deviations of the estimates in parentheses**

Species	Parameter					
	$\beta_0$ (intercept)	$\beta_1$ (%FINE)	$\beta_2$ (%SLOW)	$\beta_3$ (WIDTH)	$\beta_4$ (%UNDER)	$\beta_5$ (%CWD)
PGS	0.110 (0.298)	−0.546 (0.111)	0.166 (0.101)	0.430 (0.107)	ns	ns
TS	−3.428 (0.470)	ns	ns	−0.324 (0.194)	ns	ns
JTF	−2.384 (0.343)	−0.801 (0.172)	−0.315 (0.140)	0.398 (0.120)	−0.151 (0.135)	−0.234 (0.179)
ATF	−3.979 (0.413)	−0.460 (0.254)	ns	0.290 (0.182)	ns	ns

The corresponding covariate names are given. Note: ns denotes non-significant effects where the 90% CIs overlapped with zero.

used is represented in Fig. 1, where the variables are shown as boxes and the direction of influence is shown by arrows. This model corresponds to the final habitat models for each species (Table 1). Driving (i.e., habitat) variables were modeled as coming from a distinct probability distribution. In this study, new evidence came from the covariates being measured, or from survey information for one or more of the species. The higher-level stand- and drainage-effects were omitted from Fig. 1 since these were assigned population average values (zero) and thus were assumed to be known with certainty.

Updating is carried out using the chain rule of probability which allows the joint probability of all variables to be factored into a set of conditional distributions (Pearl, 1988). The BN structure allows this to be simplified, such that the probability of a variable  $x_j$  can be computed based only on the variables directly influencing it, called the parents of  $x_j$  ( $pa(x_j)$ ),

$$P(x_j|x_1, \dots, x_{j-1}) = P(x_j|pa(x_j))$$

The parents of each response variable in Fig. 1 have arrows pointing to that response, indicating a causal relationship. The conditional probabilities connecting the habitat variables to the responses (species probabilities) are given by the hierarchical Bayesian model Eq. (1). Methods for updating probabilities with the inclusion of new evidence are outlined in Pearl (1988) and Jensen and Nielsen (2007). We used the ‘cut’ function within WinBUGS to isolate parameter estimates fitted by the model from the modeled (missing) covariates. Otherwise the modeled values would have been included as ‘data’ and used in parameter estimation (of  $\beta$ ,  $\alpha$ , and  $\gamma$ ).

The BN modeling framework allowed us to easily model missing covariates. Furthermore, the uncertainty in the actual value of the covariates is carried forward into the prediction uncertainty. This results in predicted probabilities of occurrence with wider credibility intervals (CIs) than those predicted using an assumed (but actually unknown) value. Measured covariates were assumed known without error. Uncertainty in parameter estimates  $\beta$ ,  $\alpha$ , and  $\gamma$  (i.e., standard errors) were also incorporated into overall prediction error through their posterior distributions.

We exploited a unique property of BNs to make predictions for a species by incorporating survey information for co-occurring species. In this approach, the evidence was presence/absence survey data for one or more of the species. It was assumed that surveys did not census the species being predicted. Information is conveyed through the conditional dependence of species on a subset of the same habitat variables (model covariates). For example, PGS and TS are both influenced by WIDTH (Fig. 1), so survey information for

the presence or absence of PGS will – with WIDTH and TS unknown – alter the probability of WIDTH as being higher or lower than average. This updated (i.e., posterior) distribution for WIDTH will in turn influence the predicted probability of TS. In other words, species are linked through shared habitat variables. This information transfer only occurs when covariates are not known (modeled), otherwise species are conditionally independent given the covariates. It should be noted that model predictions are best (smallest CIs) with fully measured covariates.

A final note on terminology for those less familiar with Bayesian analyses. We used two types of prior distributions. First, the model parameters ( $\beta$ ,  $\sigma_\alpha$  and  $\sigma_\gamma$ ) were given prior distributions as are required for all Bayesian analyses. These in turn were updated using the data and Bayes theorem, resulting in the posterior distributions for the parameters (i.e., estimates) that were presented in Table 1. Second, prior distributions were used to describe the unknown (i.e., modeled) habitat covariates. Such distributions are commonly used in BNs, and in Bayesian analyses involving imputation on missing covariates (Gelman and Hill, 2007). These prior distributions were also updated to posterior distributions that described the expected distribution of the covariates given the data.

## 4. Results

The drainage and stand effects had a large influence on probability of occurrence for all species. The drainage-scale variance ( $\sigma_\gamma^2$ ) was estimated as 1.174 with 95% CI of 0.59–2.07. Predicted drainage effects were between −2.21 and 1.59 on the logit scale, across all species. Stand effects were estimated within a drainage and showed a lower variance for PGS, JTF and ATF ( $\sigma_\alpha^2$  0.45) than for TS ( $\sigma_\alpha^2$  3.91). Stand effects for TS were between −1.77 and 5.45, reflecting the large variance. TS have a patchy distribution (Welsh and Lind, 1996) which may explain the large variance of stand effects. For the other species, stand effects were between −2.5 and 2.3.

The final HB model explained 24% of the deviance over a null model, and included covariates listed in Table 1. Covariate selection was guided by 90% CIs, where variables that overlapped with zero were dropped. These variables had nominal influence on model predictions, as judged by a substantial 18 point drop in DIC between the full and reduced models. Our goal was to develop a parsimonious model that still had high predictive ability, rather than an ecological analysis on the effects of certain habitat elements on amphibians. While retaining all habitat covariates in the model was possible,

most existing habitat models have a similarly parsimonious form. Since BN predictions rely on the final model (and do not influence model selection) this parsimonious model allowed a more realistic and understandable application. The final model structure is shown graphically in Fig. 1, showing that species were influenced by common habitat covariates and were connected through these shared influences. The covariates were standardized prior to model fitting, so can be interpreted as the effect (on the logit scale) of a one standard deviation increase in measured values.

The model in this study differs from the previous analyses by Stoddard and Hayes (2005) in the use of a hierarchical structure to incorporate effects of stand and drainage effects on stream-scale predictions. For this reason, we did not expect to necessarily find all of the same significant covariates, and possibly even a change in the direction of influence for others. Such changes in inference are broadly referred to as Simpson's paradox (Pearl, 2000). For example, percentage shrub cover of the stream was shown to be important in models only considering stream-scale influences (Stoddard and Hayes, 2005). In contrast, once stand- and drainage-scale influences were accounted for, the percentage shrub cover was found to be not significant. This was due to certain drainages and stands having consistently high or low shrub cover, and the effects of shrubs could not be separated from the stand or drainage effects.

## 5. Applications

We present examples of predictions made with evidence from two sources: measured covariates and species surveys. First, model results are given with completely measured covariates to provide a reference. Second we show species predictions where all covariates were unmeasured, and contrast this with partially measured covariates. Next we incorporate survey information from co-occurring species into predictions. Finally, we incorporate a mixture of habitat and co-occurring species information. Throughout the text we referred to 'measured' covariates as one that were actually measured in the field, so they were known with certainty. We further distinguished unmeasured covariates as being 'modeled.'

### 5.1. Predictions with full information

Model predictions will have the highest precision when all covariates are measured. Predictions were made by assigning the average value to each covariate, representing full information. These predictions roughly correspond to sample-based mean occurrences of the species, except for JTF (Table 2). An exact correspondence was not expected because of the differences between model- and designed-based estimates (Thompson, 2002), but these differences were not relevant since we were concerned with point predictions rather than population estimates. The models were conditionally independent across species with all covariates measured (Fig. 1), so survey information for co-occurring species would not increase precision of the predictions.

For comparison, model predictions and standard deviations were computed where none of the covariates were

**Table 2 – Sample-based and model-based mean probability of occurrence for each species**

Species	Sample-based <sup>a</sup>	Model-based with covariates <sup>b</sup>	
		Measured	Modeled
PGS	0.53	0.527 (0.385, 0.672)	0.524 (0.198, 0.834)
TS	0.095	0.035 (0.012, 0.071)	0.037 (0.009, 0.092)
JTF	0.162	0.071 (0.035, 0.127)	0.098 (0.010, 0.337)
ATF	0.046	0.020 (0.008, 0.039)	0.024 (0.004, 0.070)

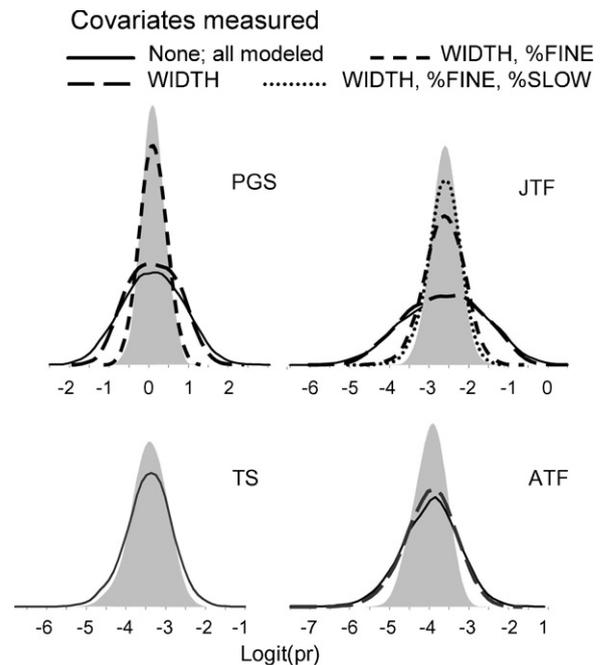
<sup>a</sup> Mean probability of occurrence across all sample points.

<sup>b</sup> Model-based predictions were made with all covariates measured or modeled (i.e., missing). 90% CIs are given in parentheses.

measured (i.e., all covariates were modeled) (Table 2), with results presented on the logit scale (Fig. 2). The substantially higher uncertainty in predictions comes from modeling the covariates. Standard deviations were roughly twice that of predictions with measured covariates, with this increase reflected in wider CI's (Table 2).

### 5.2. Prediction with missing covariates

For comparison, predictions were made where information on a subset of model covariates was available. This is a common



**Fig. 2 – Predicted probability of occurrence for each of the four species with certain habitat covariates missing. Predictions shown are posterior distributions of probability on the logit scale. Predictions with all covariates measured are shown as gray-shaded distributions. See applications for assigned covariate values. Predictions with all covariates missing (modeled) are shown as solid lines. Predictions based on one or more measured covariates are shown as dashed lines. The legend indicates which covariates were measured with the remaining covariates modeled.**

occurrence in many fields (e.g., Borsuk et al., 2004) especially where data come from a variety of sources. Stream width influenced all species and was always included as a measured covariate. In addition to WIDTH, other predictions included the measured covariates %FINE alone or in combination with %SLOW. The remaining covariates were modeled as before. All measured covariates were assumed to have mean values (all zero on the standardized scale). Results were overlaid with prediction with all covariates modeled. For reference, predictions with all covariates measured were shown in Fig. 2 as grey-shaded distributions.

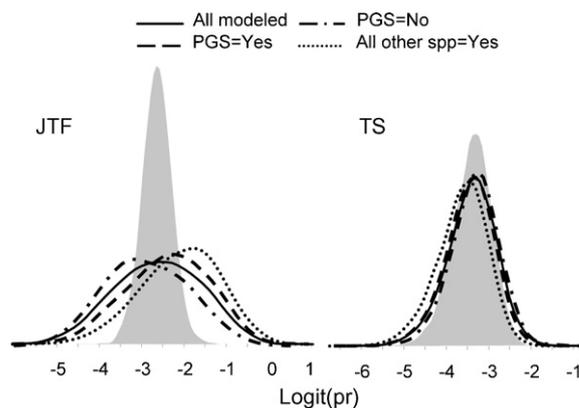
Precision of predictions was related to the amount and type of available evidence. These varied for the different species depending on the combination of measured and modeled covariates (Fig. 2). Probability (on the logit scale) of PGS was poorly predicted when only WIDTH was available, but the inclusion of %FINE decreased the prediction interval (i.e., variance of the posterior distributions) to nearly that of when all covariates were measured. In contrast to results for the other species, TS was predicted with higher error by the model (Table 2), and this was reflected in similar prediction intervals for situations where all covariates were modeled or measured (Fig. 2). Modeling missing covariates did not shift predictions higher or lower, because the assumed distributions of the covariates had the same means as the assigned values.

Basing the expected distribution of a covariate on the field surveys was reasonable in this study. The sample points were chosen randomly so we expect the sample to represent the broader study area. There was only a weak correlation between certain covariates, so no attempt was made to model these as conditional relationships. In cases where strong correlations exist, this may prove advantageous since information about one variable can provide information about another. Further, stand- and watershed timber harvesting can strongly influence habitat covariates, particularly %FINE and %CWD. As better data become available, the modeled covariate distributions should be modified to include these management effects. As the BN is structured (Fig. 1) measured information on one habitat covariate does not influence our belief about the others. This situation changes when survey information for co-occurring species is available, as demonstrated in the next section.

### 5.3. Predictions with information on co-occurring species

Survey information for co-occurring species is an additional source of information (evidence) that can be exploited by the BN framework. In this situation, survey information would be available for species that are influenced by at least one of the same covariates (shown in Table 1). That is, the model structure was conditionally dependent across species, provided at least one shared covariate was missing (Fig. 1). Predictions were made for TS and JTF with all covariates modeled and using survey information for co-occurring species.

The influence of surveyed species on the probability of non-surveyed species was entirely through the shared covariates in our model structure. For TS, this connection was weak because WIDTH – the only covariate that was shared with the other three species – had only a weak influence on these other



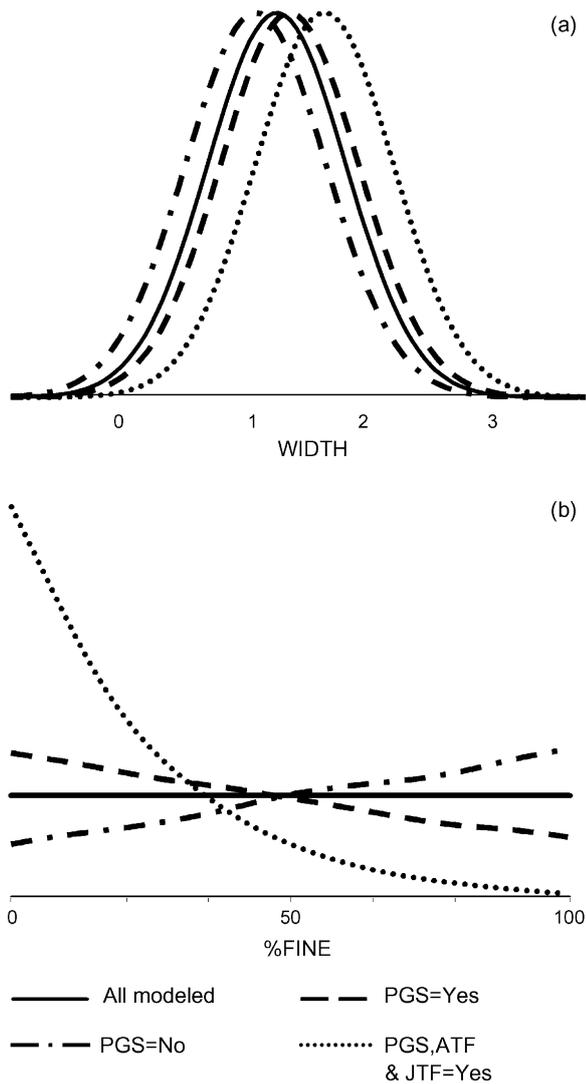
**Fig. 3 – Species occurrence probabilities for JTF and TS incorporating survey information for co-occurring species, shown on the logit scale. Predictions with all covariates measured are shown as grey-shaded distributions for reference. Predictions with all covariates modeled are shown as solid lines and incorporated no survey information. The legend indicates survey results (present = Yes, absent = No) for the co-occurring species.**

species (Table 1). With survey information on a single species, the predictions of TS was only slightly different from models with all missing covariates (Fig. 3). The reason for this can be seen in predictions for WIDTH based on the species survey information. Including survey results for PGS resulted in only a slight shift in WIDTH probability (Fig. 4a), corresponding to the direction of influence of WIDTH on PGS (Table 1). In contrast, predictions for JTF were more influenced by survey information for co-occurring species (Fig. 3). The probability of %FINE was initially modeled as being uniformly distributed over the range of observations. With the evidence that species surveys provided, the posterior distribution of %FINE was considerably altered (Fig. 4b). In this case, PGS, JTF and ATF were surveyed for and were all present, and thus the probability of %FINE was strongly expected to have a low value indicating habitats with coarse hiding cover. This in turn, altered predictions for the species that respond to %FINE.

The strength of the approach is in connecting species through shared 'driving' variables. This allows model-based predictions to incorporate other surveyed species in situations where certain covariates were missing. Predictions were best (smallest CIs) when all covariates were measured, but if measured covariates are not available this approach allows more accurate predictions than by simply assuming a mean value. As with modeling covariates, the uncertainty in the covariate value is carried forward into prediction uncertainty. There are many situations where a host of species (flora and fauna) respond to a shared set of habitat variables, and this approach allows us to learn about the variables important to their distribution and success through observed species 'responses.' This approach is not limited to presence/absence data.

### 5.4. Predictions with a mixture of information

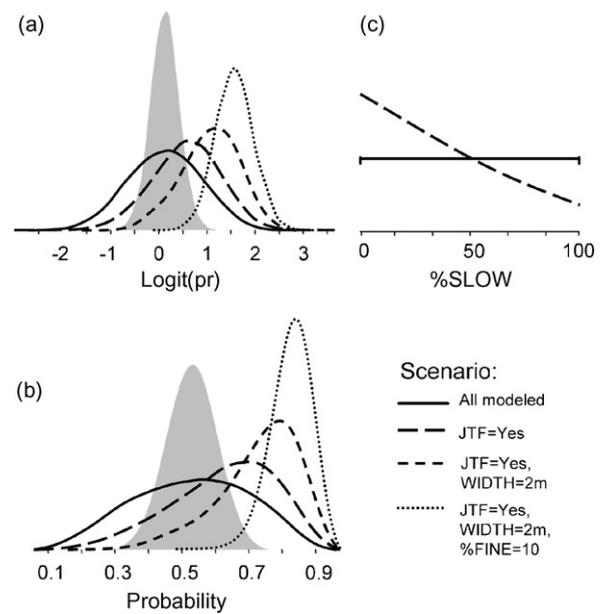
As a final example, we made predictions for a relatively abundant species PGS, with a mixture of available habitat



**Fig. 4 – Modeled and updated probabilities for habitat covariates WIDTH (a) and %FINE (b). Predictions are shown as modeled (normal or uniform priors) and updated (posterior) distributions with evidence from co-occurring species survey information, detailed in Fig. 3.**

and species information. This situation is relatively common, where available data for an area of interest collectively comes from several decades of species and habitat surveys. We also discuss possible efficiency gains in surveys by selectively modeling certain covariates, or only surveying for easily detectible and relatively common species.

For reference, predictions for PGS were contrasted with situations where all covariates were fully measured or modeled. In this simulation, WIDTH was assigned to be 2 m (1.5 S.D. > avg.) and %FINE was 10% (1.5 S.D. < avg.). Modeled covariates were incorporated into predictions in the same way as the previous examples. Incorporating positive survey results for JTF resulted in higher mean probability of occurrence for PGS (Fig. 5a and b) due to this information's influence on the modeled covariates. That is, finding JTF influenced the probability distribution of %SLOW, which was expected to have lower values (i.e., more riffles and cascades) reflecting the fact that JTF



**Fig. 5 – Probability of occurrence for Pacific giant salamander on the logit (a) and probability (b) scales, predicted with a combination of measured and modeled covariates and survey information for co-occurring species. Missing covariates were modeled as described in the text. The expected distribution for %SLOW (c) is shown as modeled (uniform prior) and updated (posterior) distributions with evidence from JTF surveys.**

was associated with faster water (Fig. 5c). Further incorporating information about WIDTH and %FINE both increased the probability of finding PGS given that these were positively and negatively related to PGS, respectively (Table 1). This example was chosen to demonstrate a relatively large shift in expected responses for a species; other responses were less dramatic but this all depended on the scenario.

## 6. Discussion

We demonstrated the utility of the BN approach for making predictions when certain habitat covariates were missing. This is a common situation in wildlife management where available data may come from different agencies, or were collected for an entirely different purpose such as stream restoration, floristic inventories or fish surveys. The ability to utilize all available datasets for answering management questions needs to be a priority, given the tight financial constraints most agencies have. The BN approach is different from a guild, or an indicator species approach to management since the different species are tied together through shared habitat variables. Predictions and inferences are unique for each species, and are related to their specific habitat needs.

Modeling unavailable covariates captures uncertainty in the actual values, and carries this uncertainty forward through the predictions (Pearl, 1988; Borsuk et al., 2004). This approach is a form of multiple imputation (Rubin, 1996), and is a better approach than making the simple assumption of a population mean value. Missing or unmeasured covariates could be

modeled using available data such as that collected during the region-wide current vegetation survey (Max et al., 1996) or local stand exams. The ability to model covariates can also allow habitat models to be used, rather than assuming an average frequency for a species. The alternative approach of re-fitting habitat models to only available information is usually untenable, especially for managers.

Including survey information for other species in predictions exploits a unique property of the BN framework (Pearl, 1988). It allows a response-mediated inference on the modeled covariates, in essence 'passing information' from one measured response variable to another through shared driving variables. This is referred to as belief updating or message passing about a modeled element (Pearl, 1988; Jensen and Nielsen, 2007). In this study, survey information on co-occurring species had little influence on the variance of predictions, but shifted the mean probability of occurrence possibly leading to different management inferences.

Our amphibian examples were somewhat contrived since surveys would routinely record all observations of the four species. However, in many situations surveys may not be possible for a complete set of species or due to pressure to reduce sampling costs. For example, 'light touch' surveys that are effective for PGS may not be for TS which often burrow into gravel in streams. Possible efficiency gains can be made using surveys that quickly record presence or absence of easily observed species (such as PGS). This type of information could also come from the Survey and Manage program on Federal land, and could then be used to help make predictions for co-occurring species. Alternatively, an efficient survey could be developed that might include songbird surveys focusing on a select species. Together with a subset of easily observed (but relevant) habitat characteristics in the field, from aerial photos or GIS information, information from these surveys could provide evidence to make precise predictions for co-occurring species. Sampling efficiency becomes critical where relatively fine-scale (stand or drainage basin) frequency predictions for species of interest are needed over large areas, such as an entire National Forest. Using the approach we describe, accurate model-based predictions could be made for a species without intensive and costly field surveys. In addition, because uncertainty in the modeled covariates and parameter estimates are carried forward in the prediction, managers can assess the precision of the predictions. If higher precision is warranted for an area, then targeted field surveys could be conducted. Predictions that account for all sources of uncertainty are needed for assessing risk (Varis and Kuikka, 1999; Borsuk et al., 2004; Lee and Irwin, 2005).

An advantage of our approach is that it does not directly rely on species–species correlations, which are often highly dependent on local conditions. This is not to say that species–species correlations are unimportant. In fact, when species are jointly dependent on a subset of shared habitat variables we would expect to find strong species correlations (negative or positive). Further, since these correlations are realized through the habitat variables our approach should lead to more robust predictions. An alternative approach includes stratifying sampling areas for one species based on the predicted frequency of a different species. Edwards et al. (2005) showed efficiency gains using habitat models for rela-

tively common lichens in the Oregon Coast Range, and then stratified the sampling design for rare lichen species based on their association with common species. Many such survey approaches are possible, but most rely on collecting new data (MacKenzie et al., 2004; Edwards et al., 2005). Our approach focused on efficiency gains from using existing datasets and survey information.

The probabilistic framework of BNs provides a natural platform for incorporating uncertainty into modeling (Clark and Gelfand, 2006). For example, the model can be expanded to incorporate species responses to partial harvesting (Corn and Bury, 1989; Olson et al., 2007), that also include uncertainty in that response. Similarly, future conditions can be modeled such as rainfall or stream levels (Borsuk et al., 2004), given appropriate information about the distribution of likely events. Simulating hypothesized changes due to climate change or management is another application. Lastly, although species detectability was not estimated during the study, this can be included in the Bayesian model (e.g., Royle and Kery, 2007). In the absence of detectability estimates, the BN framework could be used to explore the potential effects of variable detectability on population inferences. In this case, different levels of detectability (along with uncertainty in the estimates) would be modeled as a variable in the BN.

This BN approach can be conducted in tailor made commercial software such as Hugin® or Netica®, using different methods than MCMC for probability updating. Most commercial packages are also constrained to categorical variables or Gaussian distributions (see Uusitalo, 2007 and references therein). However, we wish to stress that the formal definition of BNs does not require categorical (i.e., discrete) variables (Jensen and Nielsen, 2007). Updating methods are different with continuous variables, but these can allow for richer application of these methods to common ecological problems.

The BN predictions in this study could have been made separately to the parameter estimation, but they were done simultaneously in this study for analytical convenience. In the separate case, the model structure would be the same (hierarchical logistic model) but parameter estimates would be included as new variables. The parameters ( $\beta$ 's) are then modeled as  $\sim$ MVN( $\mu$ ,  $\Sigma$ ) with terms estimated by the posterior means and variance–covariance matrix from the model fit. Alternatively, the regression equations could be fitted with ordinary least squares (Borsuk et al., 2004). BN predictions and belief updating would then be made using MCMC within WinBUGS, in a manner similar to this study. The added advantage in using a joint Bayesian estimation and prediction approach in WinBUGS or any MCMC Bayesian package is that new survey information can be used to update habitat models (Clark and Gelfand, 2006). This is a topic of ongoing study, where regional habitat models can utilize survey information at the local scale to make better local predictions.

---

## Acknowledgements

We thank M. Betts, B. McComb, V. Monleon and A. Gitelman for useful discussion on various aspects of this work. The original

amphibians survey was supported by the Cooperative Forest Ecosystem Research (CFER) program at Oregon State University, and we thank the many people who provided field assistance. This work was partially supported by USDA Forest Service, agreement 07-CR-11061800-016. This is Paper 3637 of the Forest Research Laboratory, Oregon State University, Corvallis.

## REFERENCES

- Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecol. Model.* 173, 219–239.
- Bury, R.B., Corn, P.S., 1991. Sampling methods for amphibians in streams in the Pacific Northwest. USDA Forest Service General Technical Report PNW-275. Pacific Northwest Research Station, Portland, Oregon.
- Clark, J.S., Gelfand, A.E., 2006. A future for models and data in environmental science. *Trends Ecol. Evol.* 21, 375–380.
- Corn, P.S., Bury, R.B., 1989. Logging in western Oregon: responses of headwater habitats and stream amphibians. *Forest Ecol. Manage.* 29, 39–57.
- Edwards, T.C., Cutler, D.R., Zimmermann, N.E., Geiser, L., Alegria, J., 2005. Model-based stratifications for enhancing the detection of rare ecological events. *Ecology* 86, 1081–1090.
- FEMAT (Forest Ecosystem Management Assessment Team), 1993. Forest Ecosystem Management: An Ecological, Economic, and Social Assessment. US Department of Agriculture, Portland, OR.
- Franklin, J.F., Dyrness, C.T., 1973. The natural vegetation of Washington and Oregon. USDA Forest Service General Technical Report PNW-GTR-8. Pacific Northwest Research Station, Portland, Oregon.
- Gelman, A., Hill, J.L., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge, 625 pp.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd ed. Chapman & Hall, 668 pp.
- Jensen, F.V., Nielsen, T.D., 2007. Bayesian Networks and Decision Graphs, 2nd ed. Springer, 447 pp.
- Lee, D.C., Irwin, L.L., 2005. Assessing risks to spotted owls from forest thinning in fire-adapted forests of the western United States. *Forest Ecol. Manage.* 211, 191–209.
- Lee, D.C., Rieman, B.E., 1997. Population viability assessment of salmonids by using probabilistic networks. *North Am. J. Fish. Manage.* 17, 1144–1157.
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* 10, 325–337.
- MacKenzie, D.I., Bailey, L.L., Nichols, J.D., 2004. Investigating species co-occurrence patterns when species are detected imperfectly. *J. Anim. Ecol.* 73, 546–555.
- Marsh, D.M., Trenham, P.C., 2001. Metapopulation dynamics and amphibian conservation. *Conserv. Biol.* 15, 40–49.
- Max, T.A., Schreuder, H.T., Hazard, J.W., Teply, J., Alegria, J., 1996. The region 6 vegetation inventory and monitoring system. General Technical Report PNW-RP-493, USDA Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.
- Molina, R., Marcot, B.G., Leshner, R., 2006. Protecting rare, old-growth, forest-associated species under the survey and manage program guidelines of the northwest forest plan. *Conserv. Biol.* 20, 306–318.
- Muths, E., et al., 2006. The amphibians research and monitoring initiative (ARMI): 5-year report. U.S. Geological Survey Scientific Investigations Report 2006–5224, 77 pp.
- Olson, D.H. (Ed.), 1999. Survey protocols for amphibians under the Survey and Manage provision of the Northwest Forest Plan, version 3.0. Interagency publication of the Regional Ecosystem Office, Portland, OR. BLM Publication BLM/OR/WA/PT-00/033.1792; US Government Printing Office:2000-589-124/04022, Region No.10, 310 pp.
- Olson, D.H., Anderson, P.D., Frissell, C.A., Welsh Jr., H.H., Bradford, D.F., 2007. Biodiversity management approaches for stream-riparian areas: perspectives for Pacific Northwest headwater forests, microclimates, and amphibians. *Forest Ecol. Manage.* 246, 81–107.
- Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, 2nd printing. Morgan Kaufmann, 552 pp.
- Pearl, J., 2000. Causality: Models, Reasoning and Inference. Cambridge University Press, 400 pp.
- Reckhow, K.H., 1999. Water quality prediction and probability network models. *Can. J. Fish. Aquat. Sci.* 56, 1150–1158.
- Royle, J.A., Kery, M., 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology* 88, 1813–1823.
- Rubin, D.B., 1996. Multiple imputation after 18+ years. *J. Am. Stat. Assoc.* 91, 473–489.
- Stoddard, M.A., Hayes, J.P., 2005. The influence of forest management on headwater stream amphibians at multiple spatial scales. *Ecol. Appl.* 15, 811–823.
- Thompson, S.K., 2002. Sampling. John Wiley & Sons, 400 pp.
- Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling. *Ecol. Model.* 203, 312–318.
- Varis, O., Kuikka, S., 1999. Learning Bayesian decision analysis by doing: lessons from environmental and natural resources management. *Ecol. Model.* 119, 177–195.
- Vesely, D.G., McComb, W.C., 2002. Salamander abundance and amphibian species richness in riparian buffer strips in the Oregon Coast Range. *Forest Sci.* 48, 291–297.
- Welsh Jr., H.H., Lind, A.J., 1996. Habitat correlates of the southern torrent salamander, *Rhyacotriton variegatus* (Caudata: Rhyacotritonidae), in northwestern California. *J. Herpetol.* 30, 385–398.
- Wilkins, R.N., Peterson, N.P., 2000. Factors related to amphibian occurrence and abundance in headwater streams draining second-growth Douglas-fir forests in southwestern Washington. *Forest Ecol. Manage.* 139, 79–91.