esa ECOSPHERE

Mapping marine habitat suitability and uncertainty of Bayesian networks: a case study using Pacific benthic macrofauna

Andrea Havron,[1,5,†] Chris Goldfinger,[1] Sarah Henkel,[2] Bruce G. Marcot,[3] Chris Romsos,[1] and Lisa Gilbane[4]

[1]*Active Tectonics and Seafloor Mapping Laboratory, College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 104 CEOAS Administration Building, Corvallis, Oregon 97331 USA*
[2]*Benthic Ecology Laboratory, Department of Integrative Biology, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Dr, Newport, Oregon 97365 USA*
[3]*Department of Agriculture, Pacific Northwest Research Station, United States Forest Service, 1220 SW 3rd Ave, Portland, Oregon 97204 USA*
[4]*Bureau of Ocean Energy Management, U.S. Department of the Interior, 760 Paseo Camarillo, Suite 102 (CM 102), Camarillo, California 93010 USA*

Citation: Havron, A., C. Goldfinger, S. Henkel, B. G. Marcot, C. Romsos, and L. Gilbane. 2017. Mapping marine habitat suitability and uncertainty of Bayesian networks: a case study using Pacific benthic macrofauna. Ecosphere 8(7):e01859. 10.1002/ecs2.1859

Abstract. Resource managers increasingly use habitat suitability map products to inform risk management and policy decisions. Modeling habitat suitability of data-poor species over large areas requires careful attention to assumptions and limitations. Resulting habitat suitability maps can harbor uncertainties from data collection and modeling processes; yet these limitations are not always transparent to resource managers, who increasingly rely on maps for spatial planning and risk assessment purposes. Interpretation of habitat suitability maps can be improved by visually communicating model uncertainty and data foundations. We applied Bayesian networks (BNs) to a small, marine dataset to model the probability of occurrence (PO) of benthic macrofauna. We also used BNs to create maps displaying model parameter uncertainty and data limitations. We developed BN models for three macrofauna species: a marine gastropod, *Aystris gausapata*, a marine bivalve, *Axinopsida serricata*, and a marine worm, *Sternaspis fossor*. We produced three map products from the BN models of each species: (1) a habitat suitability map of the PO projected from regional predictor variables; (2) an uncertainty map, displaying statistical variance of model predictions of occurrence probability; and (3) an experience map, displaying the empirical basis for PO predictions (equivalent sample size). Map results showed occurrence probability to be high and widespread for *Ax. serricata*, low to moderate and more limited to deeper offshore areas for *Ay. gausapata*, and low to high in shallow sandy regions and deeper silty regions, respectively, for *S. fossor*. The uncertainty and experience maps for each species helped identify regions to prioritize for future sampling. Our results are the first to show that BNs can effectively model habitat suitability of benthic macrofauna, and our detailed methods can be applied to a variety of taxa and systems. Visually describing statistical model uncertainty and equivalent sample size in map format improves interpretation of habitat suitability map predictions and supports place-based risk management of marine management.

Key words: *Axinopsida serricata; Aystris gausapata;* Bayesian network; benthic macrofauna; continental shelf; equivalent sample size; experience map; habitat suitability; Netica; probability of occurrence; *Sternaspis fossor;* uncertainty map.

Received 26 July 2016; revised 12 January 2017; accepted 22 March 2017; final version received 19 May 2017. Corresponding Editor: Hunter S. Lenihan.

Copyright: © 2017 Havron et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

[5] Present address: Department of Statistics, University of Auckland, Auckland, New Zealand.
† **E-mail:** andrea.havron@auckland.ac.nz

Introduction

The high-energy wave environment of coastal waters off Washington and Oregon is currently being considered for renewable energy development (Parkinson et al. 2015, Reikard et al. 2015). The installation of renewable energy devices and the presence of anchors maintained on the seafloor are expected to alter sedimentation patterns (Amoudry et al. 2009, Neill et al. 2009, Coates et al. 2012). While it is unknown how renewable energy devices will affect benthic macrofauna, it is anticipated they will disturb communities, either through direct (e.g., anchor attachment, cable laying) or through indirect (e.g., changes to the local current and sediment patterns, acoustic and electromagnetic effects) mechanisms (Boehlert and Gill 2010, Miller et al. 2013). Therefore, a preliminary assessment of macrofauna suitable habitat prior to the installation of multiple renewable energy devices will inform management concerned with minimizing effects on species while simultaneously planning for future development scenarios.

Habitat suitability probability modeling (HSP) is founded in Hutchinson's (1957) and Whittaker's (1960) theories of the ecological niche and environmental gradients. Habitat suitability probability modeling entails statistical analyses to extrapolate probabilities of a species' occurrence from known correlations between environmental conditions and species presence and absence or abundance (Hirzel and Guisan 2002, Franklin 2009). In the context of this study, high probabilities in habitat suitability refer to "suitable" environmental conditions that are statistically correlated with species presence and low probabilities refer to "unsuitable" environmental conditions that are highly correlated with species absence. Resulting HSP maps visualize the species' probability of occurrence (PO) across the landscape to aid in resource management. Habitat suitability probability modeling based on sampling is advantageous for producing habitat maps needed for management when limitations of cost or time prevent a complete census. Such savings are even more critical in data-poor systems, such as the marine benthos.

Multiple statistical techniques have been developed for HSP (Guisan and Zimmermann 2000, Franklin 2009). The best modeling approach provides an acceptable level of prediction accuracy given uncertainty and constraints of data as well as enough flexibility to accommodate spatially patchy and variable data distributions. Understanding the limitations and assumptions underlying statistical modeling approaches is essential for selecting the most appropriate modeling tool and for communicating model results.

Multicollinearity is a limitation of most traditional multivariate modeling frameworks (Graham 2003, Ahmadi-Nedushan et al. 2006), particularly with most principal-component and related data-reduction methods (Heckerman 1995). Multicollinearity leads to inaccurate model parameterization, decreased statistical power, exclusion of significant predictor variables (Graham 2003), and inclusion of superfluous variables. Multicollinearity needs to be considered when modeling marine benthic biota, because environmental variables used to predict species' habitats often are highly correlated in the marine benthic environment (Snelgrove and Butman 1994).

Challenges associated with benthic data collection include missing data, equipment failure and calibration error, missed samples due to weather conditions, and limited availability of continuous environmental data. In most "frequentist" statistical data-reduction methods, if a location is missing the measurement of even one covariate, the location record is omitted. Likewise, if the regional environmental dataset is missing a single covariate, then a prediction cannot be made for this location; otherwise, the model will result in inaccurate predictions (Thuiller et al. 2004). Removing the covariate with the missing information from the analysis is disadvantageous, however, if it is an important predictor for a given species.

Samples from marine benthic systems are associated with different types of uncertainty. Sources of prediction uncertainty in an ecological study can be categorized into parameter uncertainty, model structure uncertainty, inherent system variability, and measurement error, such as observational error and experimental uncertainty or error (Guisan and Thuiller 2005, Ioannidis 2005, Cressie et al. 2009, Franklin 2009). Visualizing the spatial distribution of uncertainty inherent in habitat suitability modeling improves user understanding of the limitations of the resulting habitat maps (Barry and Elith 2006, Rocchini et al. 2011).

Habitat suitability modeling has played an increasingly important role in environmental management such as for managing invasive species, reserve design, and climate change impact assessment (Guisan and Thuiller 2005, Franklin 2009, Lele et al. 2013). An appropriate assessment of uncertainty can help managers understand the limitations behind maps and models used to inform management decisions and strategies, particularly the spatial distribution of variability and error, thus providing for more scientifically informed decisions (Cleaves 1995, Barry and Elith 2006, Rocchini et al. 2011). Uncertainties naturally inherent in the modeling process, however, often go unreported (Rocchini et al. 2011). Attempts have been made to incorporate uncertainty into HSP mapping (Elith et al. 2002, Fotheringham et al. 2002, Johnson and Gillingham 2004, Bierman et al. 2010), yet these cases remain the exception. We assert that the best approach is to create additional, supporting map products to clearly communicate the spatial distribution of uncertainty and the empirical basis of habitat suitability maps.

We built Bayesian networks (BN) using Netica 5.12 (Norsys Software Corp 2014), a directed acyclic graphical modeling tool that applies Bayes' theorem to a network of variables linked by probabilities (Marcot 2006, Jensen and Nielson 2007). Bayesian inference provides an easy framework for learning (Gelman et al. 2014); habitat suitability models can be updated as new data are collected, thereby improving model performance over time. Bayesian networks remain robust to small datasets, multicollinearity, and missing data (Heckerman 1995, Kontkanen et al. 1997, Myllymäki et al. 2002, Uusitalo 2007). Bayesian networks are also designed to track and propagate uncertainty through the system (Sivia and Skilling 2006, Uusitalo 2007, Gelman et al. 2014) and, in the context of this report, can provide a final habitat suitability map along with maps visualizing prediction uncertainty and the empirical bases for model outcomes, represented here as equivalent sample size. The former map is a measure of parameter uncertainty and the latter communicates a form of experimental uncertainty. These two maps can help managers to interpret the distribution of habitat for spatial planning, thereby improving knowledge of uncertainty and data limitations underlying PO projections. We present a case study using marine macrofauna found within soft sediment substrata along the Pacific Northwest continental shelf of the western United States. These habitat suitability maps are the first of their kind for the reported species. Further, our techniques have a broader applicability across other systems and species.

Methods

Benthic macrofauna data

Motivated by plans to develop renewable energy along the Pacific Northwest continental shelf and slope, the U.S. Department of the Interior Bureau of Ocean Energy Management implemented a project to characterize benthic environments at a regional scale along the continental shelf (Goldfinger et al. 2014, Henkel et al. 2014). This project required the collection of samples across a range of locations along the continental shelf from southern Washington to northern California. We collected a box core grab sample of bottom sediment at 153 stations across eight sites of the continental shelf (Fig. 1). We retained a sub-sample of sediment to determine grain size, percent silt, and percent sand (using laser diffraction particle size analysis) as well as percent organic carbon and nitrogen (using acid composition). We filtered the remaining sediment from each grab sample through a 1.0-mm screen and all benthic macrofauna left behind were preserved for identification (Henkel et al. 2014). We recorded species observed within a sample as "present"; otherwise, we assigned the sample an "absent" value. We consider the absence data to be highly accurate since macrofauna species (animals greater in size than 1.0 mm) were the focal group for models and we thoroughly examined all samples on the 1.0-mm screen for remaining organisms. Although there is a slight chance the species was in the sub-sample retained for sediment analysis resulting in a false absence recording, this was unlikely due to the small fraction of the sample required to analyze sediment. CTD (conductivity, temperature, depth, plus additional sensors) casts were conducted at each box core sampling station to analyze temperature, dissolved oxygen, pH, fluorescence, and turbidity. See Henkel et al. (2014) for detailed methods.

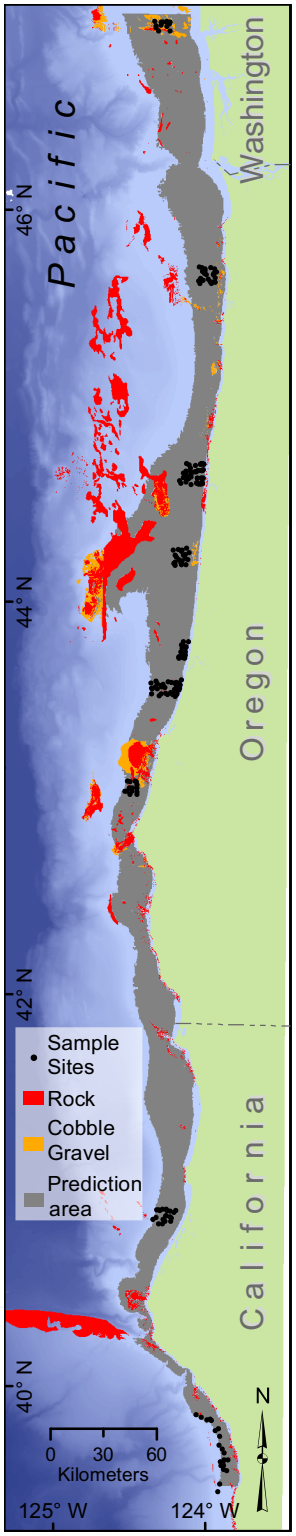


Fig. 1. Study area. Southern and northern bounds defined by macrofauna sampling locations (39°30′25.668″ to 47°01′2.64″). Eastern and western bounds defined by shallowest and deepest samples (−20 to −130 m). Areas of hard rock (red), cobble and gravel (orange) are masked from the final habitat suitability map.

We selected seven benthic macrofauna species for modeling habitat suitability based on a Primer SIMPER analysis (Clarke 1993) and for species whose variations in densities highly contributed to distinctions between assemblages at multiple sampling sites (see Henkel et al. 2014 for full description of analysis methods). Three of these seven modeled species are highlighted in this report as a case study of modeling methods: a marine gastropod, *Aystris gausapata* (Gould, 1850), a marine bivalve, *Axinopsida serricata* (Carpenter, 1864), and a marine polychaete worm, *Sternaspis fossor* (Stimpson, 1854; Appendix S1: Fig. S1). *Aystris gausapata* and *S. fossor* were selected as model candidates because their distributions were expected to change as a result of sediment changes from offshore wave energy installations. *Axinopsida serricata* was selected due to unique characteristics of its distribution, warranting further investigation with habitat suitability models to help determine the utility of the tool across a spectrum of species. The three modeled species also serve to highlight aspects and strengths of the BN modeling method.

Environmental data

We initially considered multiple sediment characteristics and water column properties for habitat suitability modeling (Table 1). A preliminary exploration of variables using frequency histograms indicated low variability in salinity (range: 33.1–33.9 PSU) and temperature (range: 7.3–9.8°C). We judged that such variability is unlikely to be biologically significant, so we excluded these two variables from habitat modeling. Further, we excluded pH, fluorescence, and turbidity due to unacceptable measurement errors likely encountered during the equipment calibration process. We derived high-resolution bathymetry data for each sampling location by using existing raster data (Goldfinger et al. 2014) in ArcGIS 10.1 (ESRI, Redlands, California, USA). We calculated an additional parameter, distance from each benthic sample station to shore in Geographic

Table 1. Description of environmental variables.

Variables	Data source	Model parameterization	Model prediction	Units
Mean grain size	Collected in situ Regional raster†	In situ	Regional raster—250 m	phi
Latitude	Collected in situ	In situ	Regional raster—250 m	degrees
Percent silt	Collected in situ	In situ	Predicted in network	percent
Percent sand	Collected in situ	In situ	Predicted in network	percent
Total organic carbon	Collected in situ	In situ	Predicted in network	percent by weight
Total nitrogen	Collected in situ	In situ	Predicted in network	percent by weight
Salinity	Collected in situ	Excluded—Insignificant	Insignificant variable	PSU
Temperature	Collected in situ	Excluded—Insignificant	Insignificant variable	°C
pH	Collected in situ	Excluded—Insufficient data	Insufficient data	pH
Fluorescence	Collected in situ	Excluded—Insufficient data	Insufficient data	μg/L
Turbidity	Collected in situ	Excluded—Insufficient data	Insufficient data	NTU
Depth	High-resolution raster†	High-resolution raster	Regional Raster—250 m	m
Distance to shore	ArcGIS raster	ArcGIS raster	Regional Raster—250 m	m

Notes: Data were either collected in situ with species or calculated as a raster in ArcGIS. Models were parameterized with in situ variables and high-resolution rasters. Variables not included in the model were excluded either due to being insignificant or having insufficient data. Variables used for prediction were all generalized to a 250-m cell size or predicted within the network.

† Goldfinger et al. (2014).

Information System using Euclidean distance analysis on a polyline shoreline.

We joined spatial data on presence and absence of benthic macrofauna species to the raster data on high-resolution bathymetry and distance to shore. Results were data on species presence, absence, and environmental variables at each sample location, including the remotely sensed variables depth and distance to shore and the in situ substrate variables mean grain size (MGS), total organic carbon (TOC), total nitrogen (TN), percent silt, and percent sand. We then constructed the BN models of habitat suitability for each of the three species using this full dataset.

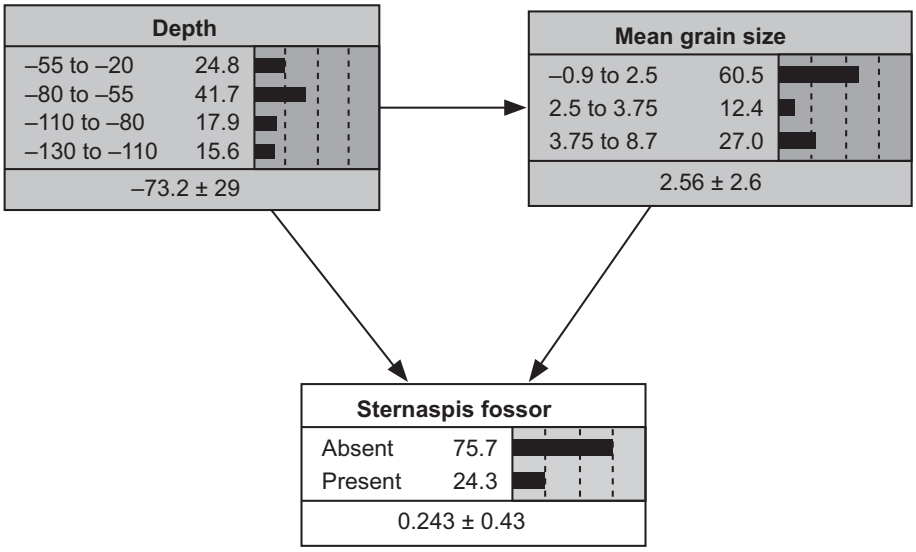
We extracted depth, distance to shore, and MGS values to a 250-m resolution point grid determined by the coarsest GIS layer, MGS (Goldfinger et al. 2014). This point grid provided continuous coverage of the region, or the extent of the study area, and was used as input in the BN model to create final predictive layers.

Model development

We designed the benthic macrofauna BN models along guidelines of Marcot et al. (2006) and Uusitalo (2007). The modeling steps included (1) variable discretization, (2) network structure development, (3) model parameterization and calibration, and (4) model prediction, selection, and validation.

Variable discretization.—Variable discretization entailed identifying quantitative ranges of state values of continuous variables (Fig. 2). Although several automated (supervised, machine-learning) techniques exist for variable discretization, none is specifically standard in the context of ecological datasets. Myllymäki et al. (2002) recommended methods that use ecologically significant breakpoints and that minimize the number of states so that each interval contains enough data to parameterize and run the model.

We used an expert-driven technique by visually inspecting frequency-value histograms from the data comparing presence and absence of each species in relation to values of each variable (Fig. 3). We initially estimated breakpoints by selecting values at the minimum and maximum range of a species presence response in the frequency-value histograms, or where histograms shifted in histogram density between present and absent. We structured a simple species-single environment BN model within Netica based on initial estimated breakpoints, parameterized conditional probability tables (CPTs), and calculated the resulting percent error based on a threshold probability of ≥ 0.5. We then incrementally increased or decreased the number of cutoff points and adjusted breakpoint values, re-parameterized CPTs, and recalculated percent error, so as to minimize error rates. Through this



Mean grain size	Depth	Absent	Present
–0.9 to 2.5	–55 to –20	100	1.96E-05
–0.9 to 2.5	–80 to –55	100	1.49E-05
–0.9 to 2.5	–110 to –80	99.9999	7.29E-05
–0.9 to 2.5	–130 to –110	49.0704	50.9296
2.5 to 3.75	–55 to –20	99.9995	0.0005
2.5 to 3.75	–80 to –55	60.6936	39.3064
2.5 to 3.75	–110 to –80	16.0001	83.9999
2.5 to 3.75	–130 to –110	38.2353	61.7647
3.75 to 8.7	–55 to –20	99.999	0.001
3.75 to 8.7	–80 to –55	0.000102	99.9999
3.75 to 8.7	–110 to –80	6.29927	93.7007
3.75 to 8.7	–130 to –110	62.5668	37.4332

Fig. 2. An example Bayesian network with respective conditional probability table (CPT). This model describes an example relationship between species and environment, where mean grain size (MGS; right node) is informed by depth (left node) and the macrofauna species (*Sternaspis fossor*; bottom node) is dependent on both depth and MGS. Each node is composed of states, which categorize the data into different bins. The CPT describes the probability of each child state for each possible combination of parent states.

iterative process, multiple discretization schemes were compared to optimize breakpoint locations and the number of states that resulted in lowest univariate classification error rates.

Network structure development.—In this step, we identified direct correlations or causal linkages among variables (Fig. 2). Previous benthic macrofauna BN models (Lockett 2012) have been used on the tree-augmented naïve (TAN) algorithm (Friedman et al. 1997), for example, as used by Aguilera et al. (2010) and Dlamini (2011), for supervised learning of the BN network structure. The TAN algorithm builds the model structure directly from the sample dataset. When no other information is available, this is an appropriate technique. Expert knowledge of environmental interactions, however, has shown to be more reliable than algorithmically derived structures

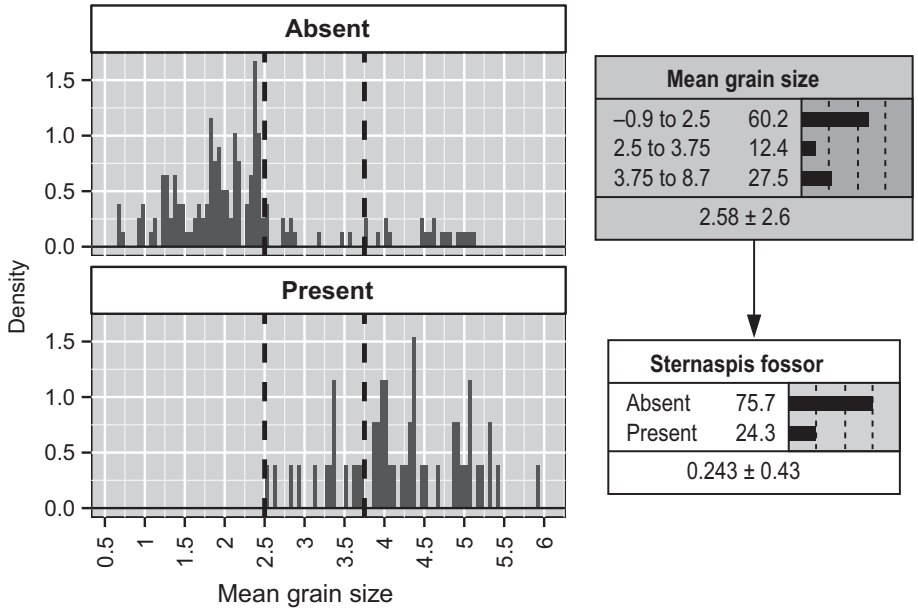


Fig. 3. Expert discretization technique to select breakpoints and state parameters for building the Bayesian network model. Breakpoints (dotted lines) are determined by visually inspecting histograms of mean grain size at sites where a species was absent (top graph) and present (bottom graph). This is an example using *Sternaspis fossor* and does not represent this species' final model.

which typically require large amounts of data to effectively learn relationships from complex systems with high variation and uncertainties (Uusitalo 2007, Alameddine et al. 2011, Chen and Pollino 2012). Further, as the TAN structure is based on the original sampling dataset alone, the incorporation of new information or data into the network requires the TAN structure to be rebuilt with each new update and can lead to significant changes in the network structure. Using an expert-defined structure based on known environmental relationships allows for the development of a more reusable network that can be updated with new information without applying significant changes to the network.

For these reasons, we used an expert-defined link structure where a combination of expert knowledge and correlations among variables within the dataset guided the design. For example, grain size, percent silt, and percent sand essentially measure the same phenomenon and therefore are highly correlated (Folk and Ward 1957). Also, TOC and TN are known to be highly correlated with percent silt (Hedges and Keil 1995). Therefore, we designed the BN models explicitly linking correlated sediment variables. Further, we calculated Pearson's correlation coefficient for all pairwise environmental variables. This step confirmed the high correlation between sediment variables and also revealed correlations between depth and the sediment variables, MGS, percent silt, percent sand, TOC, and TN (Fig. 4). Links were then added between depth and sediment variables to account for all such correlations. We chose to retain correlated variables in the model instead of omitting them because each might play different roles in predicting presence of each species.

We identified two scales of explanatory environmental variables: regional variables, or those that correspond with regional raster datasets and represent continuous coverage throughout the study area; and in situ variables, or those whose values are known only at sediment sample locations. We designed a network structure to facilitate the prediction of in situ variables by inserting intermediate nodes into the networks (Fig. 5). These intermediate nodes re-discretized regional variables (distance to shore, depth, latitude, and MGS) to best predict the correlated in situ

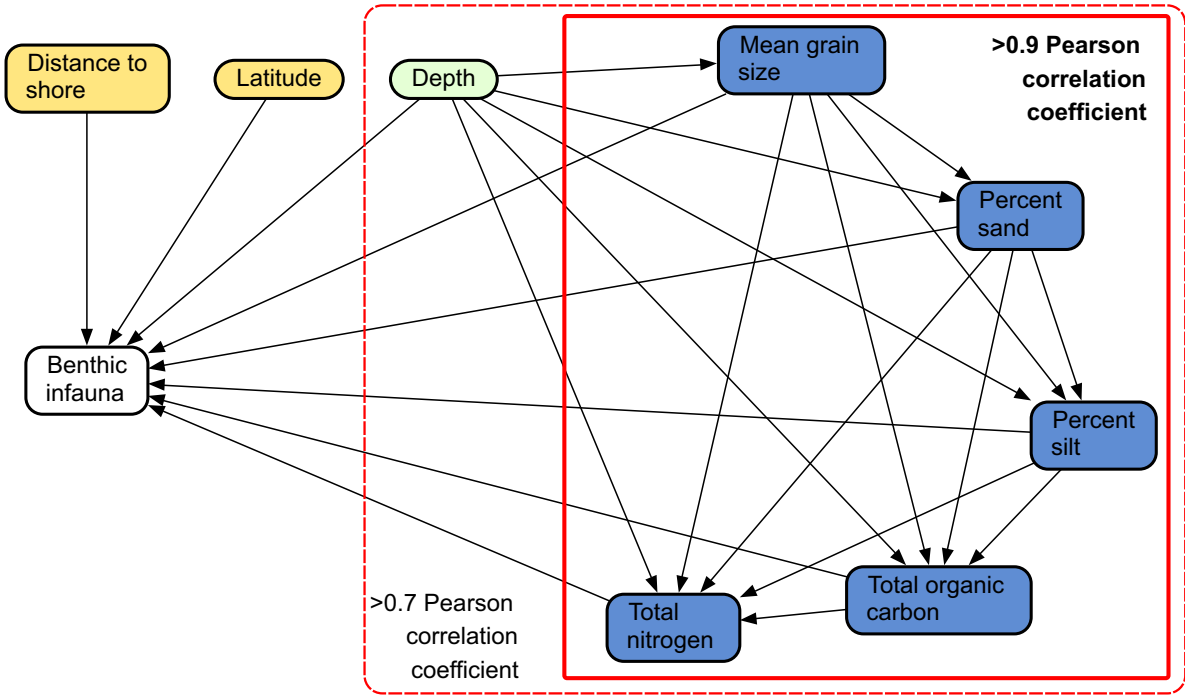


Fig. 4. Expert Bayesian network structure. The links between nodes of this structure were developed based on a prior understanding of the environment and confirmed by correlations in the data. Mean grain size, percent sand, percent silt, total organic carbon, and total nitrogen were all highly correlated with a Pearson correlation coefficient >0.9. Depth was correlated with each of the above variables with a Pearson correlation coefficient >0.7.

variables (percent silt, percent sand, TOC, and TN). Uncertainty in the projected values of the in situ variable was carried through the network into the uncertainty of the final PO prediction, expressed as the posterior probabilities of the benthic macrofauna absence and presence states.

The result of this process is a benthic macrofauna BN model framework for invertebrates living within marine sediment, which is adaptable to new species and updateable (Fig. 6). See Data S2 to download the Base Network Netica file. The BN network structure can be adapted to other benthic macrofauna species that are influenced by the suite of environmental variables presented here, with only slight modifications to species–environment discretization breakpoints. The model can be updated with new information about either a species of interest or relationships among explanatory variables (e.g., percent silt and TOC).

Model parameterization and calibration.—After models were established with uniform probability distributions, CPTs were parameterized by Netica using the expectation–maximization (EM) algorithm. The EM algorithm sequence increases the likelihood of the parameter and converges to a local maximum (Dempster et al. 1977, Wu 1983), making it a robust method for defining and updating prior and conditional probabilities, particularly when missing data are present (Watanabe and Yamaguchi 2004, Marcot 2006). We first used Netica to call the EM algorithm which parameterized the CPTs for MGS, depth, percent silt, and percent sand from the U.S. Seabed sampling database (Reid et al. 2006). This step used prior knowledge of the relationships between depth and sediment size measurements throughout the study region as a prelude to learning the species–environment relationships with the EM algorithm. The resulting, fully parameterized models with CPTs of all variables specified essentially were calibrated to the case-file datasets.

Model prediction, selection, and validation.—We built and tested 12 models for each species (Table 2). Benthic macrofauna are known to organize around substrata patterns and depth contours (Sanders 1968, Gray 1974) due to

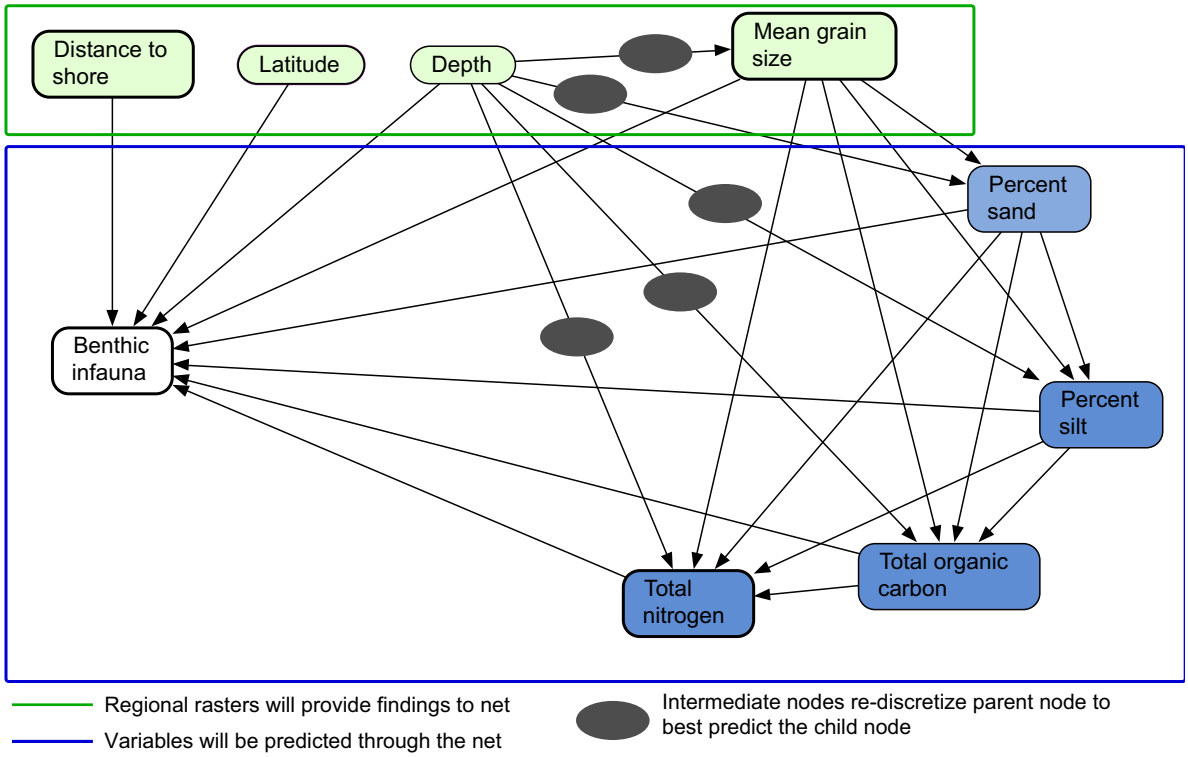


Fig. 5. Regional raster and local in situ variables. Regional raster variables predicted local in situ variables, which all combined to predict the probability of occurrence for a given species. Black circles indicate intermediate nodes, their function being to re-discretize the parent node to best predict the child node.

physiological constraints related to pressure and a reliance on a sediment-associated food supply (Snelgrove 1999). We therefore kept the variables depth and MGS in each model tested. Models varied based on the other two regional variables latitude and distance to shore, and the inclusion or exclusion of in situ sediment characteristics percent silt, percent sand, TOC, and TN. We also tested the inclusion or exclusion of intermediate nodes for predicting in situ variables. We conducted a sensitivity analysis on each network to determine the degree to which presence and absence of each species was influenced given the probability structure of the model (Marcot 2012).

We conducted a fourfold cross-validation (Stone 1974) on each candidate model and calculated performance metrics for each fold. Cross-validation measures model classification prediction accuracy and helps avoid selecting overfit models that are tightly calibrated to a dataset. We randomly split the data into four partitions. We withheld one partition for each iteration of the cross-validation fold, trained models using described methods, and made predictions for the withheld data based on the environmental raster data. We selected the threshold probability of ≥0.5 to denote between a prediction for species absence and presence. We then tallied the number of true-positive, true-negative, false-positive, and false-negative predictions for each fold in order to calculate model performance metrics. We selected final models as those with the best performance metric scores among the averaged cross-validated outcomes.

We used three model performance metrics to evaluate final models: confusion matrix error rates, spherical payoff (SP), and true skill statistic (TSS; Marcot 2012). Confusion error [0, 100%] represents the combination of type I and type II (false-positive and false-negative) error rates. Confusion error is calculated as:

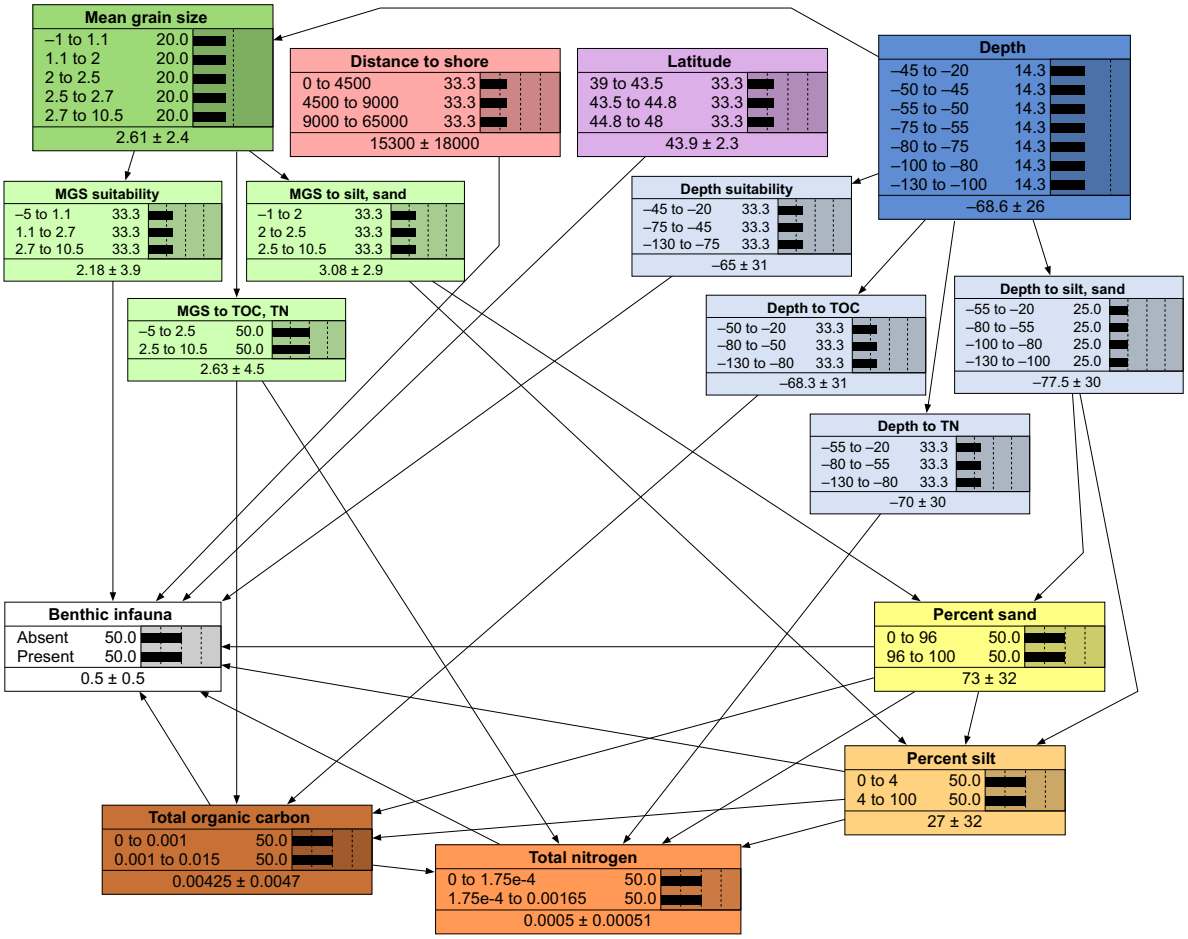


Fig. 6. Example of a re-useable and updateable Bayesian network for benthic macrofauna living within marine sediment. This framework was applied to each macrofauna species within this report prior to variable discretization and model selection. Final models can be updated with new data.

Table 2. Model tests conducted for each species.

Model no.	MGS	Depth	Latitude	Distance to shore	TOC	TN	Percent silt	Percent sand
1	X	X						
2	X	X	X					
3	X	X		X				
4	X	X	X	X				
5	X	X			X	X	X	X
6	X	X	X		X	X	X	X
7	X	X		X	X	X	X	X
8	X	X	X	X	X	X	X	X

Notes: In total, there were 12 model permutations for each species as models 5–8 were tested with and without intermediate nodes. Mean grain size (MGS) and depth were included in each model test as previous research has established their importance to benthic invertebrate distribution patterns. TOC, total organic carbon; TN, total nitrogen.

$$\text{Confusion error} = \frac{\text{FPos} + \text{FNeg}}{\text{FPos} + \text{TPos} + \text{FNeg} + \text{TNeg}}$$

where FPos = false positive, FNeg = false negative, TPos = true positive, and TNeg = true negative. A false-positive error, in the context of habitat suitability models, occurs if a model predicts that a species is present in a particular environment when, in fact, it is absent; a false-negative error occurs if a model predicts that a species is absent in a particular environment when, in fact, it is present. Best-performing models have 0% confusion error. Spherical payoff [0, 1] was chosen as it out-performs AUC, Area Under the receiver-operating characteristic Curve (Marcot 2012), where with best-performing models, SP = 1. Spherical payoff is calculated as:

$$\text{SP} = \text{MOAC} \times \frac{P_c}{\sqrt{\sum_{j=1}^{n} P_j^2}}$$

where MOAC = mean probability value of a given state averaged over all cases, P_c = the predicted probability of the correct state, P_j = the predicted probability of state j, and n = total number of states (B. Boerlage, *personal communication*). True skill statistic [−1, 1] was chosen as it is independent of prevalence and preferred over AUC for presence–absence models due to the necessity to select a threshold probability for denoting between species presence and absence during prediction (Allouche et al. 2006, Marcot 2012). A TSS score of 1 represents a model with no error, a score of 0 represents a model with random error, and a score of −1 represents a model with total error. True skill statistic is calculated as:

$$\begin{aligned}\text{TSS} &= \frac{(\text{TPos} \times \text{TNeg}) - (\text{FPos} \times \text{FNeg})}{(\text{TPos} + \text{FNeg}) \times (\text{FPos} + \text{TNeg})} \\ &= \text{sensitivity} + \text{specificity} - 1\end{aligned}$$

where FPos = false positive, FNeg = false negative, TPos = true positive, and TNeg = true negative (Allouche et al. 2006). Each performance metric has different assumptions. Confusion error is based on the highest probability state and conflates error types, which may oversimplify the utility of the model; SP is influenced by the number of states in the response variable; and TSS accounts for probabilities but also conflates error types (Marcot 2012). For these reasons, all three metrics were compared when selecting final models.

Finally, an opportunity arose during the course of this study to conduct a field validation from the Northwest National Marine Renewable Energy Center's Pacific Marine Energy Center, South Energy Test Site (SETS; Fig. 7), where tests of offshore renewable energy devices are planned. An additional 14 box core grab samples were collected in August and again in October of 2013. Whereas the data from the SETS site provided an opportunity to validate model predictions, the spatial extent of the SETS site (93 km^2) compared to the overall region of prediction

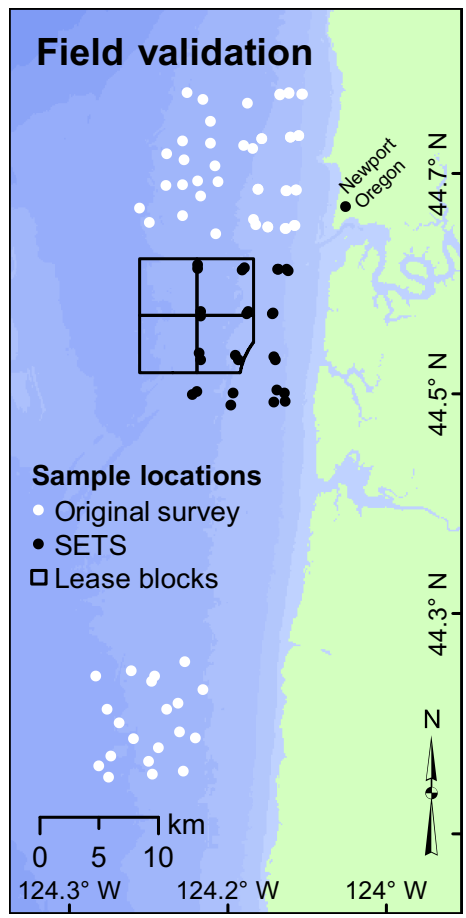


Fig. 7. Study area of field validation. South Energy Test Site (SETS) sampling stations (black dots) indicate box core samples taken on August and October 2013 and used in field validation. White dots indicated the closest sample locations sites, Newport and Cape Perpetua, used to train predictive models.

(15,880 km^2), and the small SETS sample size, prevented the data from being used exclusively to select models. Rather, validation with the SETS data was used to confirm results from the fourfold cross-validation approach and to aid in selecting between two models with close results.

We compared species presence or absence known from SETS samples with model prediction results within the SETS region. If the species was observed in a SETS sample, the model prediction at the same location was considered a true-positive result if its PO was ≥0.5, and a false-negative result (type II error) if its PO was below 0.5. If the species was absent from a SETS sample, the model prediction at the same location was considered a true-negative result if its PO was below 0.5, and a false-positive result (type I error) if its PO was above 0.5. Using this information, performance metrics were calculated.

Map creation

Applying the regional raster data as findings to the final model networks, the models calculated a posterior PO and standard error of prediction for each species at each raster cell location. From this output, we produced three maps for each species showing PO, parameter uncertainty, and equivalent sample size (referred to as "experience" in BN modeling). We used a database of regional latitude, depth, MGS, and distance to shore with a 250-m resolution cell size to generate PO predictions. Depth and MGS regional rasters were developed using the methods described by Goldfinger et al. (2014). The MGS raster, being a product of a spatial kriging analysis, had an associated mean square error of 0.8 phi. This error was included with MGS values as input into the BN when making predictions, allowing the error to be incorporated into the uncertainty of the final PO prediction. Regions of rock, cobble, and gravel were masked from the final predictive maps because the models were developed only for soft sediment species.

Uncertainty maps communicated the posterior distribution of the PO prediction, which is a measure of parameter uncertainty in the probability predictions (Marcot 2012). We created uncertainty maps by mapping the standard deviation of the expected value of the PO based on 0 = fully unsuitable environments and 1 = fully suitable environments. As the species response node follows a Bernoulli distribution, the standard deviation is: $\sqrt{PO(1-PO)}$. The species response node of our models consisted of two states: present and absent; therefore, uncertainty in probability predictions represented PO values.

Experience maps described the percentage of the benthic macrofauna sampling dataset used to inform each unique state probability. We created these maps by accessing Netica experience tables. Netica automatically updates the experience value for each unique combination of states as models are learned. If the environmental states of a CPT row are consistent with the environmental states of a specific observation from the data, the new experience number for the CPT row is increased by one. The experience value therefore reports the number of cases (observations) used by the EM learning algorithm to parameterize each row of the CPT values in the model, or the equivalent sample size for each unique combination of environmental states. We used an arcPy Con (Spatial Analyst) statement within Python (van Rossum 1991) to develop the experience maps (Data S1). Rasters for each in situ variable were first generated from the most probable outcome predicted from Netica. A conditional statement (Con) was then used to match the environmental states of a given stacked raster cell to a row in the CPT, which had an assigned experience value from the Netica experience table. This value was then assigned to the cell location of the stacked rasters. Experience values were mapped as percentage of the overall sample size, therefore visualizing the percentage of data used to learn predictions. Experience maps represent experimental uncertainty as they visualize the distribution of environmental combinations that receive more sampling effort than other combinations. High experience values occur outside sampling locations if environmental conditions correspond to those of regions heavily sampled. Experience values differ for each species, as they are dependent on the unique environmental variables important to the species of interest.

Results

The following results are reported for each species: (1) model performance metrics, (2) HSP maps, (3) uncertainty maps, (4) experience maps, and (5) SETS field validation results.

Table 3. *Axinopsida serricata* model results. A good performance score has a low confusion error rate (error rate), high SP†, and high TSS‡.

		TSS		SP		Error rate	
Models	Test	Simple	Complex	Simple	Complex	Simple	Complex
1	Fourfold cv	0.45	—	0.87	—	15	—
2	Fourfold cv	**0.51**	—	**0.85**	—	**15**	—
3	Fourfold cv	0.33	—	0.83	—	24	—
4	Fourfold cv	0.44	—	0.83	—	20	—
5	Fourfold cv	0.45	0.46	0.87	0.87	15	14
6	Fourfold cv	0.43	0.44	0.86	0.87	17	17
7	Fourfold cv	0.44	0.48	0.84	0.86	20	17
8	Fourfold cv	0.44	0.46	0.82	0.85	20	20
1	Sets	0.51	—	0.79	—	26	—
2	Sets	**0.51**	—	**0.79**	—	**26**	—
3	Sets	0.24	—	0.78	—	33	—
4	Sets	0.24	—	0.78	—	33	—
5	Sets	0.41	0.41	0.74	0.79	30	30
6	Sets	0.41	0.41	0.74	0.79	30	30
7	Sets	0.21	0.21	0.76	0.78	33	33
8	Sets	0.21	0.21	0.76	0.78	33	33

Notes: Simple models refer to ones without intermediate nodes, while complex models refer to ones with intermediate nodes. The scores from the final selected model are in boldface. SP, spherical payoff; TSS, true skill statistic.

† SP values range [0, 1], where 1 denotes best model performance.

‡ TSS values range [−1, 1], where 1 represents a perfectly performing model with no error, 0 a model with totally random error, and −1 a model with total error.

Axinopsida serricata (Carpenter, 1864) is a marine bivalve in the family Thyasiridae within the Lucinoida order. A very ubiquitous species, it was seen in 83% of 153 sampling stations. The preferred BN model (Appendix S2: Fig. S1, Data S2) selected for *Ax. serricata* from the fourfold cross-validation approach was Model 2 without intermediate nodes (Table 3). We chose this model because it had the highest TSS score. It also had one of the lowest error rates, as well as a high SP score. More emphasis was given to the TSS score because *Ax. serricata* is a highly prevalent species and TSS is independent of prevalence. The preferred model included depth, MGS, and latitude. The final model trained using all data had a TSS score of 0.58, an SP score of 0.91, and an 11% overall classification error rate. Sensitivity analysis indicated that the bivalve habitat suitability was most sensitive to MGS followed by depth and latitude (Table 4). The network conditioned on species being present (Appendix S2: Fig. S2) suggested that *Ax. serricata* was found uniformly throughout

Table 4. Sensitivity analysis results.

	Axinopsida serricata		*Aystris gausapata*		*Sternaspis fossor*	
	Model 2		Model 3		Model 6	
Rank	Variable	Variance reduction (%)	Variable	Variance reduction (%)	Variable	Variance reduction (%)
1	MGS	16.8	MGS	6.2	Silt	26.2
2	Depth	12.4	Distance to shore	3.9	MGS	25
3	Latitude	0.1	Depth	1.3	TOC	24.5
4					Sand	23.5
5					TN	23.2
6					Depth	16
7					Latitude	5.1

Notes: Percent variance reduction represents the degree to which each environmental variable is explained by the variation in the species posterior probability distributions. Variables are listed in the order of importance for each species. MGS, mean grain size; TOC, total organic carbon; TN, total nitrogen.

the latitudinal range, in slightly deeper waters and higher MGS than was typical for the region.

Aystris gausapata (Gould, 1850) is a marine gastropod in the Columbellidae family within the Neogastropoda order. This species of snail was present in 43% of 153 sampling stations. The preferred BN model (Appendix S3: Fig. S1, Data S2) selected for *Ay. gausapata* from the fourfold cross-validation approach was Model 3 without intermediate nodes (Table 5). We chose this model because it had the highest TSS score and lowest error rate. The SP score was slightly lower than several other models, yet improvements in SP were not great enough to justify choosing a model with a better SP score but worse TSS score and error rate. The preferred model included MGS, depth, and distance to shore. The final model trained using all data had a TSS score of 0.43, an SP score of 0.79, and a 29% error rate. Sensitivity analysis indicated that the snail habitat suitability was most sensitive to MGS, followed by distance to shore and depth (Table 4). The network conditioned on species being present (Appendix S3: Fig. S2) suggested that *Ay. gausapata* was found in mid-range distance to shore, mid-depth, and lower MGS than was typical for the region.

Sternaspis fossor (Stimpson, 1854) is a polychaete in the family Sternaspidae within the Terebellidae order. This species of marine worm was present in 24% of 153 sampling stations. The preferred BN model (Appendix S4: Fig. S1, Data S2) selected for *S. fossor* from the fourfold cross-validation approach was Model 6 with intermediate nodes (Table 6). We chose this model because it had the highest TSS score, one of the lowest error rates, and one of the higher SP scores. It should be noted, however, that Model 2 and Model 6, both without intermediate nodes, had the highest SP score, similar error rates, and a TSS only slightly worse than the preferred model. The preferred model was chosen over these two because the improvement in its TSS score was slightly better than the improvement of the SP score of the latter two. The preferred model included all variables except distance to shore. The final model trained using all data had a TSS score of 0.89, an SP score of 0.96, and a 5% error rate. Sensitivity analysis indicated that the polychaete habitat suitability was most sensitive to percent silt, followed by MGS, TOC, percent sand, TN, depth, and latitude (Table 4). The network conditioned on species being present (Appendix S4: Fig. S2)

Table 5. *Aystris gausapata* model results. A good performance score has a low confusion error rate (error rate), high SP†, and high TSS‡.

		TSS		SP		Error rate	
Models	Test	Simple	Complex	Simple	Complex	Simple	Complex
1	Fourfold cv	0.02	—	0.71	—	48	—
2	Fourfold cv	0.29	—	0.75	—	35	—
3	Fourfold cv	**0.36**	—	**0.72**	—	**30**	—
4	Fourfold cv	0.33	—	0.74	—	31	—
5	Fourfold cv	0.19	0.15	0.71	0.7	39	43
6	Fourfold cv	0.29	0.15	0.74	0.73	35	43
7	Fourfold cv	0.24	0.17	0.71	0.7	37	40
8	Fourfold cv	0.33	0.23	0.75	0.74	33	37
1	Sets	0.48	—	0.77	—	26	—
2	Sets	0.11	—	0.74	—	33	—
3	Sets	**0.25**	—	**0.74**	—	**41**	—
4	Sets	0.41	—	0.73	—	48	—
5	Sets	0.03	0.4	0.74	0.74	30	30
6	Sets	0.35	0.03	0.72	0.71	44	48
7	Sets	0.17	0.09	0.72	0.68	37	44
8	Sets	0.48	−0.05	0.69	0.67	48	52

Notes: Simple models refer to ones without intermediate nodes, while complex models refer to ones with intermediate nodes. The scores from the final selected model are in boldface. SP, spherical payoff; TSS, true skill statistic.

† SP values range [0, 1], where 1 denotes best model performance.

‡ TSS values range [−1, 1], where 1 represents a perfectly performing model with no error, 0 a model with totally random error, and −1 a model with total error.

Table 6. *Sternaspis fossor* model results. A good performance score has a low confusion error rate (error rate), high SP†, and high TSS‡.

		TSS		SP		Error rate	
Models	Test	Simple	Complex	Simple	Complex	Simple	Complex
1	Fourfold cv	0.65	—	0.91	—	11	—
2	Fourfold cv	0.69	—	0.92	—	11	—
3	Fourfold cv	0.56	—	0.89	—	15	—
4	Fourfold cv	0.63	—	0.92	—	11	—
5	Fourfold cv	0.65	0.58	0.91	0.89	11	12
6	Fourfold cv	0.68	**0.72**	0.92	**0.91**	12	**11**
7	Fourfold cv	0.56	0.56	0.9	0.89	15	14
8	Fourfold cv	0.62	0.65	0.91	0.91	12	12
1	Sets	1	—	1	—	0	0
2	Sets	1	—	1	—	0	0
3	Sets	1	—	1	—	0	0
4	Sets	1	—	1	—	0	0
5	Sets	1	1	1	1	0	0
6	Sets	1	**1**	1	**1**	0	**0**
7	Sets	1	1	1	1	0	0
8	Sets	1	1	1	1	0	0

Notes: Simple models refer to ones without intermediate nodes, while complex models refer to ones with intermediate nodes. The scores from the final selected model are in boldface. SP, spherical payoff; TSS, true skill statistic.

† SP values range [0, 1], where 1 denotes best model performance.

‡ TSS values range [−1, 1], where 1 represents a perfectly performing model with no error, 0 a model with totally random error, and −1 a model with total error.

suggested that *S. fossor* was found in lower to mid-latitudes, deeper water, higher MGS, TOC and TN, higher percent silt, and lower percent sand than was typical for the region.

The HSP map of *Ax. serricata* (Fig. 8a) depicts high PO throughout most of the region with small pockets of moderately unsuitable environments found near shore in shallow regions. A PO value of exactly 0.5 was made for 0% of the prediction area, and a PO value between 0.49 and 0.51 was made for 0.3% of the prediction area. The uncertainty map (Fig. 8b) reflects the patterns in the HSP maps: Regions of high PO correspond with higher precision (low uncertainty) around the posterior probability, whereas areas of the map with predicted probability around 0.5 correspond with regions of lower precision (high uncertainty). Highest uncertainty occurs in the nearshore, shallow regions, particularly off the mouth of the Columbia River between Oregon and Washington, a unique sedimentary environment which was not sampled in this study. The experience map (Fig. 8c) indicates that the largest percentage of data informed probabilities from the southern, deep region, followed by the shallow to mid-depth, mid-latitude region, and with the least amount of data informing probabilities in the northern, deep region, and in the shallow to mid-depth regions of the most northern and southern extents. Predictions learned from less than one observation (0 experience) accounted for 5.9% of the prediction area.

Field validation analysis at the SETS site (Fig. 8d) indicated that *Ax. serricata* was more frequently present at deeper stations and more frequently absent at shallow stations, following similar patterns of PO predictions. Seven model predictions were in error (Fig. 8d) compared to 28 SETS observations: five type I errors (BN model predicted presence but the species was absent in SETS sample data) and two type II errors (BN model predicted absence but the species was present in SETS sample data). Three of the five type I errors occurred at stations that sampled species presence one season and species absence another season. This temporal difference in sampling effort may indicate a region that has variability in suitable conditions, but further temporal sampling would need to occur in order to model this trend. The remaining model prediction errors occurred on the boundary between suitable and unsuitable environments.

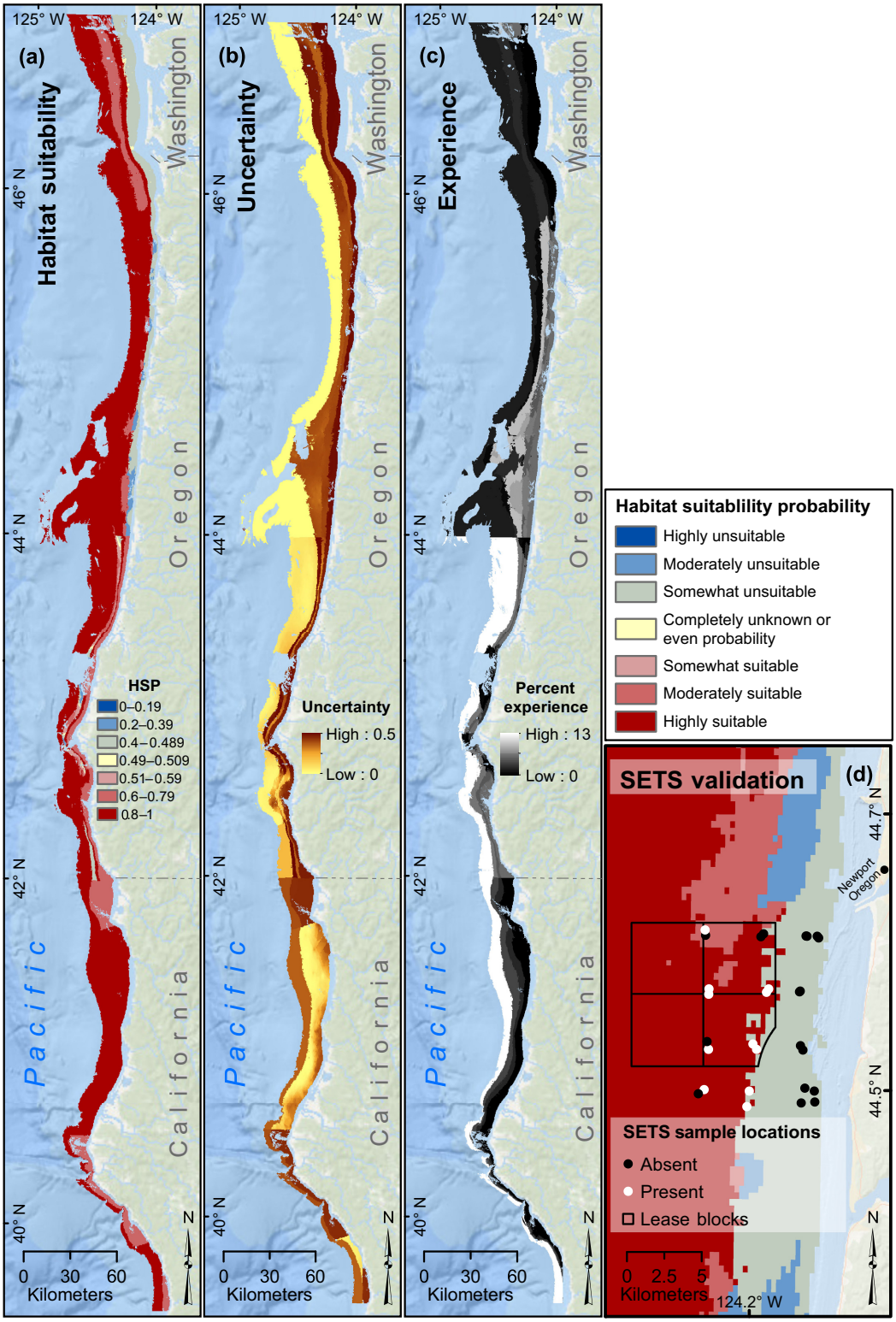


Fig. 8. *Axinopsida serricata*. Habitat suitability (a) ranges from 0 (blue: very unlikely probability of suitable

(Fig. 8. *Continued*)
environments) to 0.5 (yellow: unknown/even probability) to 1 (red: very likely probability of suitable environments). Uncertainty (b) ranges from 0 (light: high precision and low uncertainty in probability estimate) to 0.5 (dark: low precision and high uncertainty in probability estimate). Experience (c) ranges from 1 (light: high percentage of data informed probabilities) to 0 (dark: low percentage of data informed probabilities). South Energy Test Site (SETS) (d) sample site locations of observed species presence/absence over habitat suitability probability model predictions.

The HSP map of *Ay. gausapata* (Fig. 9a) depicts areas of low PO in deeper offshore areas. In addition, large regions are classified as somewhat unsuitable environments (PO between 0.4 and 0.6), including the region off the Columbia River mouth. A PO value of exactly 0.5 was made for 3.5% of the prediction area, and a PO value between 0.49 and 0.51 was made for 7.6% of the prediction area. The uncertainty map (Fig. 9b) expresses low precision (high uncertainty) throughout most of the region. Lowest uncertainty values are found in the deeper, offshore areas, where the species is predicted to be absent. The experience map (Fig. 9c) indicates small pockets in the shallow, mid-latitude regions where probabilities were informed by a higher percentage of data. In general, the majority of predictions were informed by little data. Predictions learned from less than one observation (0 experience) accounted for 19.4% of the prediction area.

Field validation analysis at the SETS site (Fig. 9d) indicated that *Ay. gausapata* was more frequently present at deeper stations and more frequently absent at shallower stations, although there was slightly greater error between observations and PO patterns than with the other species. Eleven model predictions were in error (Fig. 9d) compared to 28 SETS observations: three type I errors (BN model predicted presence but the species was absent in SETS sample data) and eight type II errors (BN model predicted absence but the species was present in SETS sample data). Six of the type II errors occurred in areas predicted to be somewhat unsuitable (PO: ~0.47). The remaining two type II errors occurred in areas predicted to have moderately unsuitable environments (which we defined as PO: 0.2–0.39). One type I error occurred in an area predicted to have somewhat suitable environments (PO: 0.51–0.59), and the other two occurred in areas predicted to have moderately suitable environments (PO: 0.6–0.79).

Overall, metric results from cross-validation and field validation (Table 5) indicate this model to have poorer performance than other species modeled. Corresponding maps help to communicate this error, uncertainty, and lack of data informing the model.

The HSP map of *S. fossor* (Fig. 10a) depicts low PO throughout shallow, sandy regions, transitioning to high PO in deeper, silty regions. A PO value of exactly 0.5 was made for 4.6% of the prediction area, and a PO value between 0.49 and 0.51 was made for 10.9% of the prediction area. The uncertainty map (Fig. 10b) follows the patterns of the HSP maps: Regions of highly suitable or highly unsuitable environments correspond with higher precision (low uncertainty) around the posterior probability, whereas areas of the map with intermediate probabilities of suitable environments correspond with regions of lower precision (high uncertainty). Areas where PO equal 0.5 occur in the southern, nearshore regions and in northern deeper regions. The nearshore environment in the south expresses unique, unsampled conditions, where the shelf drops off steeply close to shore. The experience map (Fig. 10c) indicates that the largest percentage of data informed probabilities associated with the mid-latitude, shallow, sandy habitat regions, and much of the remaining area corresponded with little to no experience. Higher probabilities of occurrence were informed by a lower percentage of data. This is likely because this less sampled species occupies a more specialized niche, representing an under-sampled combination of environmental parameters within this study. Predictions learned from less than one observation (0 experience) accounted for 30.7% of the prediction area.

Field validation analysis at the SETS site (Fig. 10d) indicated that *S. fossor* was absent in all SETS samples, corresponding with a consistent prediction of absence throughout the SETS

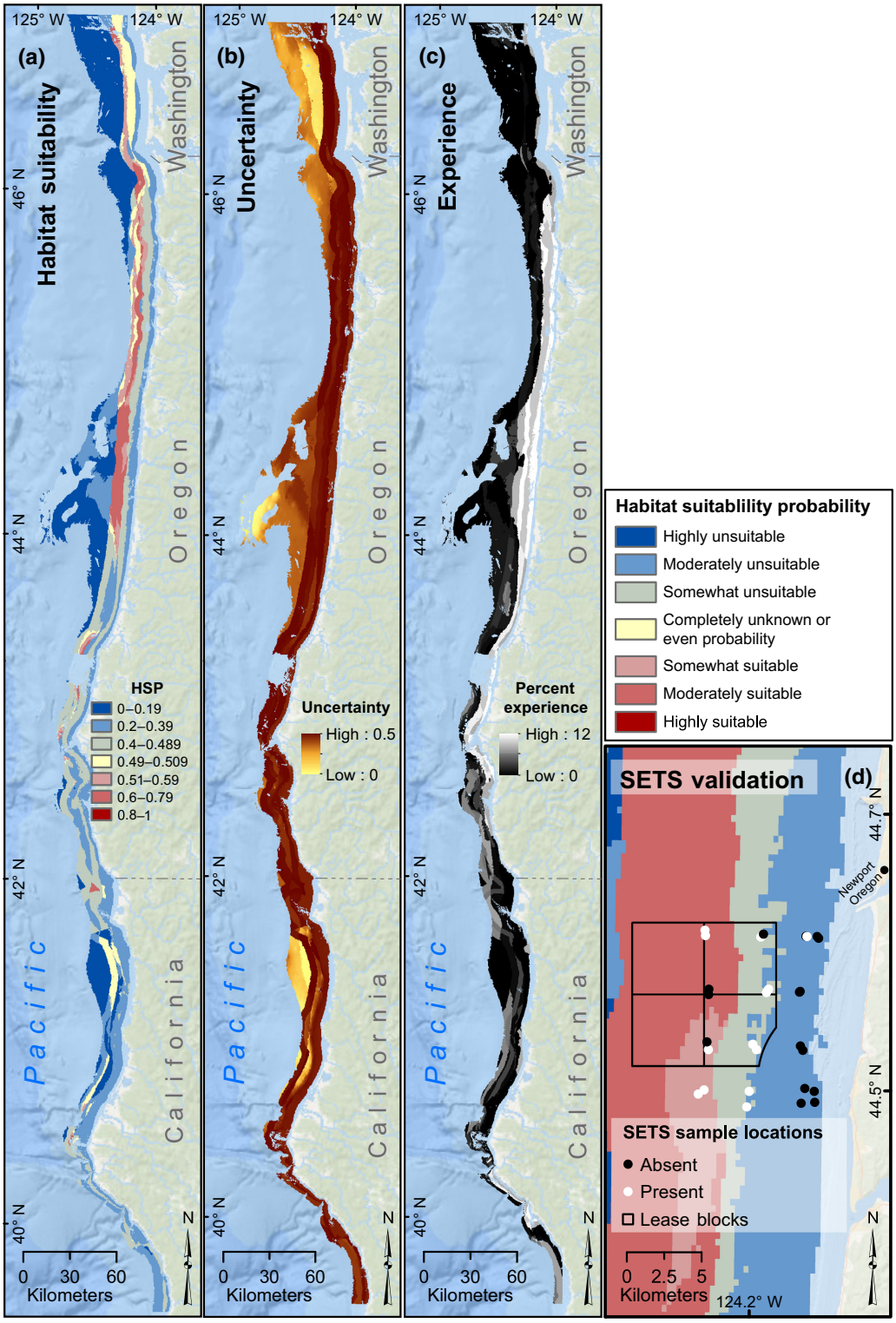


Fig. 9. *Aystris gausapata.* Habitat suitability (a) ranges from 0 (blue: very unlikely probability of suitable

(Fig. 9. *Continued*)
environments) to 0.5 (yellow: unknown/even probability) to 1 (red: very likely probability of suitable environments). Uncertainty (b) ranges from 0 (light: high precision and low uncertainty in probability estimate) to 0.5 (dark: low precision and high uncertainty in probability estimate). Experience (c) ranges from 1 (light: high percentage of data informed probabilities) to 0 (dark: low percentage of data informed probabilities). South Energy Test Site (SETS) (d) sample site locations of observed species presence/absence over habitat suitability probability model predictions.

sample area. Therefore, no type I error of false presence was observed (Fig. 10d) between SETS sample observations and model predictions, although it was not possible to determine type II errors of false absence from the SETS data given the species' absence in all field samples. Overall, metric results from cross-validation and field validation (Table 6) indicate this model to have good performance.

Discussion

Knowledge of marine benthic species is limited due to costs, time, and challenges related to marine data collection. Habitat suitability predictive modeling provides a cost-effective strategy to analyze species distribution patterns by converting statistical species–environment associations to spatial maps (e.g., Smith et al. 2007). As the process is an approximation, the development of multiple maps of prediction probabilities, uncertainty, and experience to convey statistical confidence and underlying effort will improve interpretation and utility of habitat suitability maps. As maps are frequently used to make management and policy decisions affecting natural resources, understanding uncertainty and the limitations of map products will better inform decisions and strategies (e.g., Catenacci and Giupponi 2013).

Our study outlines methods to visualize different forms of uncertainty to aid in management interpretation. Bayesian networks have previously been used to develop maps visualizing uncertainty, for example, in ecosystem services (Grêt-Regamey et al. 2013, Landuyt et al. 2015), and in soil risk management (Troldborg et al. 2013), but none is known to have been used to develop maps visualizing empirical sample size, nor uncertainties resulting from a habitat suitability analysis.

The visual representation of uncertainty improves interpretability of habitat suitability maps (Stelzenmüller et al. 2010, Kininmonth et al. 2014) and provides an innovative and general way to display variability in the expected value of the posterior probability. Probability values from a binary-outcome network contain inherent information on uncertainty. High or low probabilities of occurrence indicate low uncertainty; probabilities equal to 0.5 indicate high uncertainty. We decided it still advantageous to report uncertainty map methods and results for two reasons. First, this relationship is less clear when the response node consists of more than two states; therefore, methods are important for model extensions to abundance or habitat quality. Second, uncertainty is easier to interpret when expressed as a solid color gradient (Aerts et al. 2003); therefore, the additional maps improve the communication of results to the target audience of resource managers. Additional types of uncertainty maps can be developed, such as model structure uncertainty maps using model averaging techniques which measure the variability from several different model outputs (Raftery et al. 1997).

The uncertainty and experience maps can be used in conjunction for improved interpretation of predictions. High uncertainty means, in the case of these binary-outcome models, to consult the associated experience maps to determine how much confidence to place in the outcome given the amount of supporting data. For example, a PO equal to 0.5 can mean either no data were available to update the prior probability of 0.5, or a posterior probability calculated from case data resulted in the value due to inherent uncertainty in the system. Either scenario will result in a high uncertainty. Using the experience map can help differentiate between the two possibilities, the implications of which can be quite different, representing either knowledge absence or presence.

The experience map is also a novel product that communicates the percentage of data informing probabilities in the model, or the equivalent

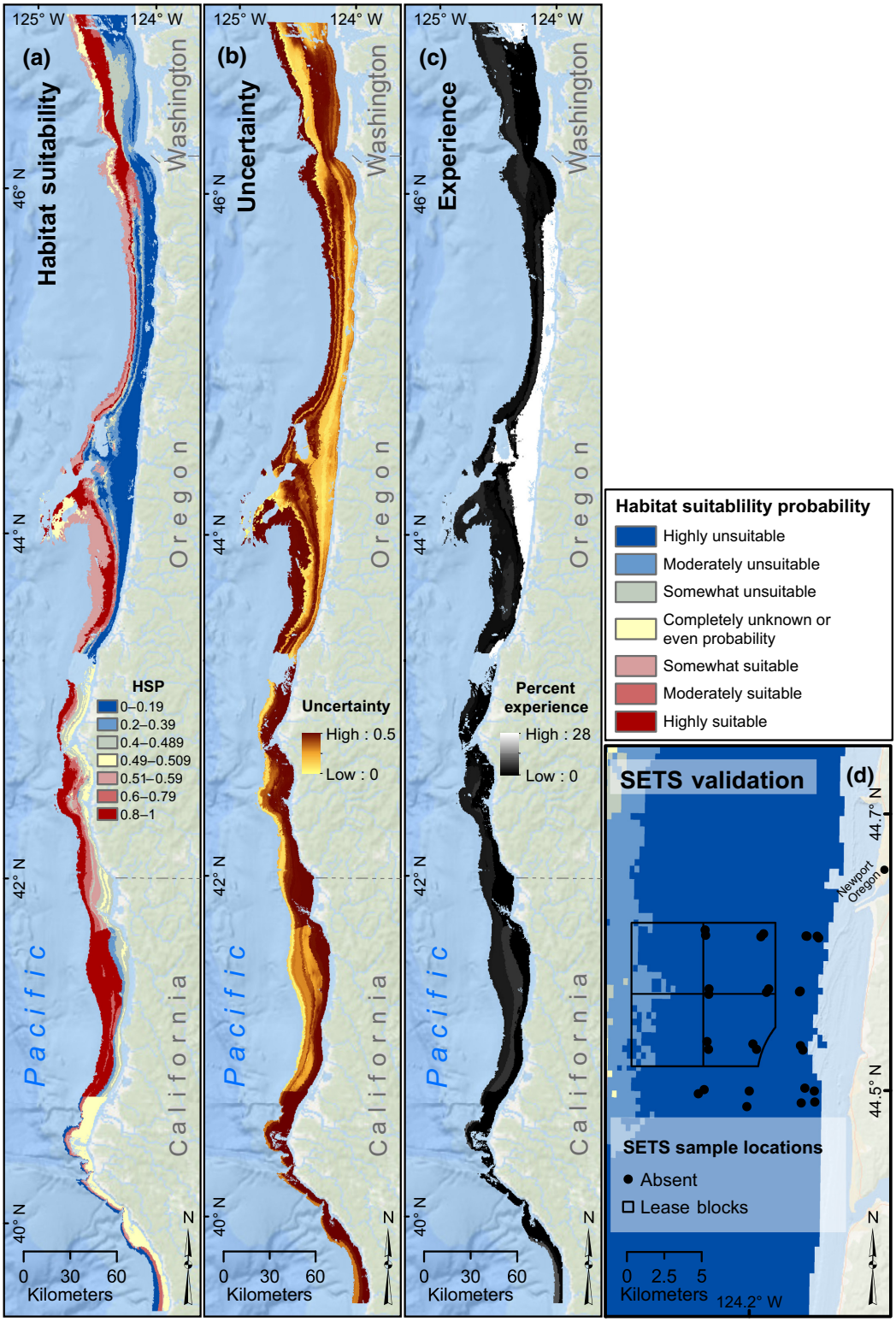


Fig. 10. *Sternaspis fossor.* Habitat suitability (a) ranges from 0 (blue: very unlikely probability of suitable

(Fig. 10. *Continued*)

environments) to 0.5 (yellow: unknown/even probability) to 1 (red: very likely probability of suitable environments). Uncertainty (b) ranges from 0 (light, high precision and low uncertainty in probability estimate) to 0.5 (dark, low precision and high uncertainty in probability estimate). Experience (c) ranges from 1 (light, high percentage of data informed probabilities) to 0 (dark, low percentage of data informed probabilities). South Energy Test Site (SETS) (d) sample site locations of observed species presence/absence over habitat suitability probability model predictions.

sample size. This map helps identify uncertainty stemming from experimental design, primarily from sampling strategy. Experience values are not necessarily correlated with uncertainty values. A region can be high in experience and low in confidence (high in uncertainty) if a large percentage of data were collected from somewhat suitable environmental conditions. In contrast, a region can have high confidence (low uncertainty) and low experience if a small percentage of data were collected from highly suitable environmental conditions. An example is the off-shore area of central Oregon where results on *Axinopsida serricata* suggested high PO, high confidence, yet very low experience (Fig. 8a–c).

Due to properties mentioned above, experience maps also provide a novel interpretation to identify regions to target for future sampling effort. Hirzel and Guisan (2002) recommended large sampling effort over both space and time for HSP modeling. However, this is a costly process in the marine environment, and therefore, directed sampling effort can maximize information while minimizing cost. They also recommended regular grid sampling over stratified sampling; however, they recognized that stratified sampling may be a necessary and effective cost-saving strategy. In the case of the latter, they recommended using environmental knowledge to stratify the sample. Experience maps presented here can be used for such purposes to design future stratified sampling effort. Species within this study will be used as an example for how the experience maps can inform future sampling effort.

For example, the *Aystris gausapata* HSP model was based on low experience throughout the study area likely due to weak statistical correlations between species presence and absence and environmental variables. Management recommendations include increased sampling effort throughout the region and identification of new environmental predictor variables. The *Sternaspis fossor* HSP model was based on higher experience in the mid-latitude, shallow, sandy portion of the study area yet had higher PO values in deeper, silty environments which were also associated with lower experience, indicating a sampling bias toward unsuitable environmental conditions. A management recommendation would be to increase sampling effort in deeper, silty regions to increase information about this species. The *Ax. serricata* HSP model was based on higher experience in offshore, deeper water in the south compared with the north. While this species was highly prevalent throughout the region, model predictions would be improved by increasing sampling effort in the northern, offshore region.

Future efforts should endeavor to increase sampling across consistent depth ranges along the latitudinal gradient. Current datasets entailed in samples spatially clustered by latitude necessitated a simple discretization structure of latitude in the models. Such simplification may not have captured biologically significant latitudinal patterns. Further, the simple discretization structure resulted in boundary artifacts, as appeared in *Ax. serricata's* experience and uncertainty maps around the 44th parallel (Fig. 8b, c).

Due to the lack of uniformity in sampling effort, under-sampled regions in this study include the shallowest and deepest extents, the southern portion of the study site where deeper, siltier environments are found closer to shore, and the northern region surrounding the output of the Columbia River. The Columbia River plume is a large driver in sedimentary patterns in the near-shore environment and can influence benthic conditions across a broad latitudinal range (Wright and Nittrouer 1995). A better understanding of this system will improve model predictions, which were predominantly reported for this part of the study area as

somewhat suitable environments, high uncertainty, and low experience for all species.

Finally, further field validation studies should be conducted to assess model prediction accuracy. The usefulness of the field validation at the SETS site was constrained by its sample size and scope as well as its sampling within areas of moderate to high experience. Maps indicated that modeling results from the SETS sampling area had a moderate level of experience with environmental features important for predicting *Ay. gausapata* and *Ax. serricata*, and high levels with environmental features important for predicting *S. fossor.* To improve the models, new sampling sites should be identified to maximize geographic coverage and unique environmental areas. Additional field validation, over both time and space, with subsequent model updating will improve model prediction success (Aguilera et al. 2011).

The experience map could be improved in future iterations by including the uncertainty associated with MGS raster generated from a kriging process (0.8 phi). This uncertainty was propagated through the network when calculating posterior probabilities and sometimes resulted in MGS probabilities spread over several states. In the case where the expected MGS value resulted in a 0.5 probability prediction, yet uncertainty induced the MGS probability to be spread over multiple states, the probability prediction could possibly be greater or less than 0.5 even when the experience value equaled 0. Such an event could also occur as a result of missing variable information, in which case the BN model reverted to using default prior probability distributions to calculate the posterior probabilities.

Our methods for developing and accessing multiple models for each species, and parameterizing the probability values in the network using the EM algorithm, were highly useful and could be applied to other studies for evaluating different species. We tested models based on a threshold probability of 0.5, above which corresponds to a prediction of presence and below, a prediction of absence. A lower threshold may be preferred when modeling sensitive species to minimize false-negative error and would likely have implications for model selection and field validation.

Habitat suitability probability maps developed in this study reflect static probability predictions of occurrence for benthic macrofauna species and not abundance per se, given regional raster information. These models are useful despite being limited by incomplete knowledge of individual species–environment associations (Tantipisanuh et al. 2014, Hamilton et al. 2015). Our models are predominantly learned from species relationships to geomorphic variables that are easily measured on a regional scale, allowing for the development of predictive maps across a large study area. Models, however, do not capture local variability associated with biological, chemical, or demographic dynamics, nor do they make a prediction of habitat quality or a species' equilibrium distributions. Our models also do not address spatial autocorrelation, which has been shown to bias habitat suitability models (Dormann 2007). Species–environment associations with small-scale features and latent variables can be inserted into the BN, allowing for future model improvements as new knowledge of these relationships becomes available. In addition, species counts from a wider sampling effort can inform on species abundance or ordinal habitat preferences, which could enhance confidence in predictions.

Acknowledgments

This material is based on work supported by the U.S. Department of the Interior, Bureau of Ocean Energy Management, Environmental Studies Program under Cooperative Agreement M10AC20002. We thank the captains and crews of the R/V Pacific Storm, R/V Elakha, Miss Linda, Derek M. Baylis, Marine Applied Research and Exploration, David Evans, and Associates for assistance in data collection. We thank Kristin Politano for the processing of invertebrate and sediment samples and Daniel Lockett for assistance in model performance tests during model selection. We would like to thank Rebecca Flitcroft and two anonymous reviewers for feedback that improved the quality of the manuscript.

Literature Cited

Aerts, J. C. J. H., K. C. Clarke, and A. D. Keuper. 2003. Testing popular visualization techniques for representing model uncertainty. Cartography and Geographic Information Science 30:249–261.

Aguilera, P. A., A. Fernández, R. Fernández, R. Rumí, and A. Salmerón. 2011. Bayesian networks in environmental modelling. Environmental Modelling and Software 26:1376–1388.

Aguilera, P., A. Fernandez, F. Reche, and R. Rumi. 2010. Hybrid Bayesian network classifiers: application to species distribution models. Environmental Modelling and Software 25:1630–1639.

Ahmadi-Nedushan, B., A. St-Hilaire, M. Bérubé, É. Robichaud, N. Thiémonge, and B. Bobée. 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Research and Applications 22:503–523.

Alameddine, I., Y. Cha, and K. H. Reckhow. 2011. An evaluation of automated structure learning with Bayesian networks : an application to estuarine chlorophyll dynamics. Environmental Modelling and Software 26:163–172.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43:1223–1232.

Amoudry, L., P. S. Bell, K. S. Black, R. W. Gatliff, R. Helsby, A. J. Souza, P. D. Thorne, and J. Wolf. 2009. A scoping study on: research into changes in sediment dynamics linked to marine renewable energy installations. NERC Marine Renewable Energy Theme Action Plan Report. British Geological Survey, Edinburgh, UK.

Barry, S., and J. Elith. 2006. Error and uncertainty in habitat models. Journal of Applied Ecology 43: 413–423.

Bierman, S. M., A. Butler, G. Marion, and I. Kühn. 2010. Bayesian image restoration models for combining expert knowledge on recording activity with species distribution data. Ecography 33:451–460.

Boehlert, G. W., and A. B. Gill. 2010. Environmental and ecological effects of ocean renewable energy development: a current synthesis. Oceanography 23:68–81.

Catenacci, M., and C. Giupponi. 2013. Integrated assessment of sea-level rise adaptation strategies using a Bayesian decision network approach. Environmental Modelling and Software 44:87–100.

Chen, S. H., and C. A. Pollino. 2012. Good practice in Bayesian network modelling. Environmental Modelling and Software 37:134–145.

Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117–143.

Cleaves, D. A. 1995. Assessing and communicating uncertainty in decision support systems: lessons from an ecosystem policy analysis. AI Applications 9:87–102.

Coates, D., J. Vanaverbeke, and M. Vincx. 2012. Enrichment of the soft sediment macrobenthos around a gravity based foundation on the Thorntonbank. Pages 41–54 *in* S. Degraer, R. Brabant, and B. Rumes, editors. Offshore wind farms in the Belgian part of the North Sea: heading for an understanding of environmental impacts. Royal Belgian Institute of Natural Sciences, Management Unit of the North Sea Mathematical Models, Marine Ecosystem Management Unit, Brussels, Belgium.

Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19:553–570.

Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39:1–38.

Dlamini, W. M. 2011. Bioclimatic modeling of southern African bioregions and biomes using Bayesian networks. Ecosystems 14:366–381.

Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography 16:129–138.

Elith, J., M. A. Burgman, and H. M. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecological Modelling 157:313–329.

Folk, R., and W. Ward. 1957. Brazos River Bar: a study in the significance of grain size parameters. Journal of Sedimentary Petrology 27:3–26.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, West Sussex, UK.

Franklin, J. 2009. Mapping species distributions: spatial inference and prediction. Cambridge University Press, New York, New York, USA.

Friedman, N., D. Geiger, and M. Goldszmidt. 1997. Bayesian network classifiers. Machine Learning 29:131–163.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. Bayesian data analysis. Third edition. CRC Press, Boca Raton, Florida, USA.

Goldfinger, C., S. K. Henkel, C. Romsos, A. Havron, and B. Black. 2014. Benthic habitat characterization offshore the Pacific Northwest, volume 1: evaluation of continental shelf geology. OCS Study BOEM 2014-662. U.S. Department of the Interior, Bureau of Ocean Energy Management, Pacific OCS Region, Camarillo, California, USA.

Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84:2809–2815.

Gray, J. S. 1974. Animal-sediment relationships. Oceanography and Marine Biology: An Annual Review 12:223–261.

Grêt-Regamey, A., S. H. Brunner, J. Altwegg, M. Christen, and P. Bebi. 2013. Integrating expert knowledge into mapping ecosystem services tradeoffs for sustainable forest management. Ecology and Society 18:1–34.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

Hamilton, S. H., C. A. Pollino, and A. J. Jakeman. 2015. Habitat suitability modelling of rare species using Bayesian networks: model evaluation under limited data. Ecological Modelling 299:64–78.

Heckerman, D. 1995. A tutorial on learning with Bayesian networks. Technical report MSR-TR-95-06. Microsoft Research, Redmond, Washington, USA.

Hedges, J. I., and R. G. Keil. 1995. Sedimentary organic matter preservation: an assessment and speculative synthesis. Marine Chemistry 49:81–115.

Henkel, S. K., C. Goldfinger, C. Romsos, L. G. Hemery, A. Havron, and K. Politano. 2014. Benthic habitat characterization offshore the Pacific Northwest, volume 2: evaluation of benthic communities. OCS Study BOEM 2014-662. U.S. Department of the Interior, Bureau of Ocean Energy Management, Pacific OCS Region, Camarillo, California, USA.

Hirzel, A., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling 157:331–341.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415–427.

Ioannidis, J. P. A. 2005. Why most published research findings are false. PLoS Medicine 2:e124.

Jensen, F. V., and T. D. Nielson. 2007. Bayesian networks and decision graphs. Second edition. Springer, New York, New York, USA.

Johnson, C. J., and M. P. Gillingham. 2004. Mapping uncertainty: sensitivity of wildlife habitat ratings to expert opinion. Journal of Applied Ecology 41: 1032–1041.

Kininmonth, S., S. Lemm, C. Malone, and T. Hatley. 2014. Spatial vulnerability assessment of anchor damage within the Great Barrier Reef World Heritage Area, Australia. Ocean & Coastal Management 100:20–31.

Kontkanen, P., P. Myllymaki, T. Silander, H. Tirri, and P. Grunwald. 1997. Comparing predictive inference methods for discrete domains. Pages 311–318 *in* P. Smyth and D. Madigan, editors. Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics, Fort Lauderdale, Florida, USA, January 4–7, 1997. Society for Artificial Intelligence and Statistics, New Jersey, USA.

Landuyt, D., K. Van der Biest, S. Broekx, J. Staes, P. Meire, and P. L. M. Goethals. 2015. A GIS plug-in for Bayesian belief networks: towards a transparent software framework to assess and visualise uncertainties in ecosystem service mapping. Environmental Modelling and Software 71:30–38.

Lele, S. R., E. H. Merrill, J. Keim, and M. S. Boyce. 2013. Selection, use, choice and occupancy: clarifying concepts in resource selection studies. Journal of Animal Ecology 82:1183–1191.

Lockett, D. E. 2012. A Bayesian approach to habitat suitability prediction. Thesis. Oregon State University, Corvallis, Oregon, USA.

Marcot, B. G. 2006. Characterizing species at risk I: modeling rare species under the Northwest Forest Plan. Ecology and Society 11:1–10.

Marcot, B. G. 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecological Modelling 230:50–62.

Marcot, B. G., J. D. Steventon, G. D. Sutherland, and R. K. McCann. 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Canadian Journal of Forest Research 36:3063–3074.

Miller, R. G., Z. L. Hutchison, A. K. Macleod, M. T. Burrows, E. J. Cook, K. S. Last, and B. Wilson. 2013. Marine renewable energy development: assessing the benthic footprint at multiple scales. Frontiers in Ecology and the Environment 11:433–440.

Myllymäki, P., T. Silander, H. Tirri, and P. Uronen. 2002. B-Course: a web-based tool for Bayesian and causal data analysis. International Journal on Artificial Intelligence Tools 11:369–387.

Neill, S. P., E. J. Litt, S. J. Couch, and A. G. Davies. 2009. The impact of tidal stream turbines on large-scale sediment dynamics. Renewable Energy 34: 2803–2812.

Norsys Software. 2014. Netica. Version 5.12. Norsys Software, Vancouver, British Columbia, Canada. http://www.norsys.com/

Parkinson, S. C., K. Dragoon, G. Reikard, G. García-Medina, H. T. Özkan-Haller, and T. K. A. Brekken. 2015. Integrating ocean wave energy at large-scales: a study of the US Pacific Northwest. Renewable Energy 76:551–559.

Raftery, A. E., D. Madigan, and J. A. Hoeting. 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association 92:179–191.

Reid, J. A., J. M. Reid, C. J. Jenkins, M. Zimmermann, S. J. Williams, and M. E. Field. 2006. usSeabed: Pacific Coast (California, Oregon, Washington)

offshore surficial-sediment data release. U.S. Geological Survey Data Series 182. Version 1.0. https://coastalmap.marine.usgs.gov/flexweb/westcoast/pacificcoast/

Reikard, G., B. Robertson, and J.-R. Bidlot. 2015. Combining wave energy with wind and solar: short-term forecasting. Renewable Energy 81:442–456.

Rocchini, D., J. Hortal, S. Lengyel, J. M. Lobo, A. Jiménez-Valverde, C. Ricotta, G. Bacaro, and A. Chiarucci. 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Progress in Physical Geography 35:211–226.

Sanders, H. L. 1968. Marine benthic diversity: a comparative study. American Naturalist 102:243–282.

Sivia, D., and J. Skilling. 2006. Data analysis: a Bayesian tutorial. Oxford University Press, New York, New York, USA.

Smith, C. S., A. L. Howes, B. Price, and C. A. McAlpine. 2007. Using a Bayesian belief network to predict suitable habitat of an endangered mammal: the Julia Creek dunnart (*Sminthopsis douglasi*). Biological Conservation 139:333–347.

Snelgrove, P. V. R. 1999. Getting to the bottom of marine biodiversity: sedimentary habitats. BioScience 49:129–138.

Snelgrove, P. V. R., and C. A. Butman. 1994. Animal-sediment relationships revisited: cause versus effect. Oceanography and Marine Biology: An Annual Review 32:111–177.

Stelzenmüller, V., J. Lee, E. Garnacho, and S. I. Rogers. 2010. Assessment of a Bayesian Belief Network-GIS framework as a practical tool to support marine planning. Marine Pollution Bulletin 60:1743–1754.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society. Series B (Methodological) 36: 111–147.

Tantipisanuh, N., G. A. Gale, and C. Pollino. 2014. Bayesian networks for habitat suitability modeling: a potential tool for conservation planning with scarce resources. Ecological Applications 24: 1705–1718.

Thuiller, W., M. B. Araújo, and S. Lavorel. 2004. Do we need land-cover data to model species distributions in Europe? Journal of Biogeography 31:353–361.

Troldborg, M., I. Aalders, W. Towers, P. D. Hallett, B. M. McKenzie, A. G. Bengough, A. Lilly, B. C. Ball, and R. L. Hough. 2013. Application of Bayesian Belief Networks to quantify and map areas at risk to soil threats: using soil compaction as an example. Soil & Tillage Research 132:56–68.

Uusitalo, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling 203:312–318.

van Rossum, G. 1991. Python. Version 2.7.2. Python Software Foundation, Wilmington, Delaware, USA. https://www.python.org

Watanabe, M., and K. Yamaguchi. 2004. Incomplete data and the EM algorithm. Pages 9–20 *in* M. Watanabe and K. Yamaguchi, editors. The EM algorithm and related statistical models. Marcel Dekker, New York, New York, USA.

Whittaker, R. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30:279–338.

Wright, L. D., and C. A. Nittrouer. 1995. Dispersal of river sediments in coastal seas: six contrasting cases. Estuaries 18:494–508.

Wu, C. F. J. 1983. On the convergence properties of the EM algorithm. Annals of Statistics 11:95–103.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1859/full