==========

ELEMENT DISTRIBUTION MODELING: A PRIMER

- or -

MUSINGS UPON DEPICTING THE GEOGRAPHIC ARRANGEMENT OF BIOTA, WITH AMPLE REFERENCE TO THE HARSH LESSONS OF THE DARK MASTER EXPERIENCE

VERSION 2.0 - 1 June 2006

Prepared by: Gary P. Beauvais [1]
Douglas A. Keinath [1]
Pilar Hernandez [2]
Larry Master [3]
Rob Thurston [1]

[1] Wyoming Natural Diversity Database, University of Wyoming, 1000 E. University Ave // Department 3381, Laramie, Wyoming, USA, 82071; phone 307 766-3023; fax 307 766-3026; beauvais@uwyo.edu / dkeinath@uwyo.edu / thurston@uwyo.edu

[2] Department of Zoology / Landscape Ecology Laboratory, University of Toronto, 25 Harbord St., Toronto, Ontario, Canada, M5S 3G5; phone 416 946-7886; fax 416 978-8532; pilar_hernandez@natureserve.org

[3] Zoology Department, NatureServe, 11 Avenue de Lafayette, 5th Floor, Boston, MA, USA; phone 617 542-1908 x230; fax 617-542-1620; larry_master@natureserve.org

==========

1.0 INTRODUCTION

This document is intended to introduce element distribution modeling (EDM) to an audience composed primarily of state Natural Heritage programs. It should be relevant to a more general audience as well, as it assumes only that the reader is a trained biologist with experience in managing field data on the occurrence of free-ranging organisms or their habitats (i.e., ecological communities), and has a working knowledge of statistics and geographic information systems (GIS).

This document is not an exhaustive discussion of all aspects of EDM, which would require at least a textbook-sized treatment. It takes an explicitly practical focus - how to produce reliable predictive distribution maps using technology and techniques now available to, or efficiently developed by, state Natural Heritage program personnel and other professionals working in applied natural resource science. As with most scientific pursuits there is a lot of action at the theoretical frontier of EDM. Some is very interesting, some will certainly lead to profoundly better applications someday, and some is discussed herein. But the main objective here is facilitating the application of EDM *today*. The reader is directed to the literature cited section of this document, and in particular to the publications listed in Table 1, for more detailed discussions of EDM theory, history, and cutting-edge techniques.

Table 1. Some publications that discuss conceptual details and theory relevant to element distribution modeling, listed chronologically.

Guisan and Zimmermann (2000)
Pulliam (2000)
Corsi et al. (2001)
Ferrier et al. (2002)
Scott et al. (2002)
Elith and Burgman (2003)
Rushton et al. (2004)
Guisan and Thuiller (2005)
Elith et al. (2006)

Be aware that EDM goes by a few other names: predictive distribution modeling, predictive range mapping, species distribution mapping, habitat distribution mapping, ecological niche modeling, and various other combinations of these and similar terms. This diversity of terms indicates just how fast the field is evolving. Many laboratories and research initiatives are committing to EDM, and new techniques and research results are being presented, literally, every month. The 2004 Annual Meeting of the Society for Conservation Biology established a separate EDM symposium just to corral the large number of papers submitted on the topic. This frenzy of activity can intimidate the scientist trying to become familiar enough with EDM to apply it in a practical way. Hopefully this paper organizes some of the confusion, and presents enough background to allow interested parties to quickly and effectively join the fray.

Readers are encouraged to become familiar with the terms in the GLOSSARY before proceeding through this document. There are fine-points of certain terms, such as "range" vs. "distribution" vs. "habitat", that are important to keep in mind.

This is version 2.0 of a live document, and is an update of Beauvais et al. (2004). All comments/ additions/ deletions/ corrections/ suggestions are not only welcomed but encouraged. Once a critical mass of changes has been accumulated the document will be updated to the next version and re-distributed, thus keeping at least some pace with relevant advances in EDM. Refer to Section 6.0 for a discussion of an EDM modeling tool currently under development that will remain similarly live - that is, it will not only help state Natural Heritage programs and similar users implement EDM as it currently stands, but will also be periodically updated and re-distributed with practical advances in EDM technology to keep such users near the forefront of the field.

2.0 WHAT IS ELEMENT DISTRIBUTION MODELING?

EDM, as discussed here and in the context of the basic mission of state Natural Heritage programs, is a process that maps environments predicted to be suitable for occupation (and, conversely, environments predicted to be unsuitable for occupation) by a given element in a given area. Depending upon the details of the procedure the final model may map an area in gradations of suitability, rather than simple suitable/ unsuitable terms.

This awkward definition is commonly abbreviated to "habitat model", "distribution model", or even "range map" in the literature, but it is important to recognize important differences in terminology. EDM does not directly model habitat, because a strict definition of habitat includes considerations of reproduction and survival as well as simple occupation (Morrison et al. 1992). Nor does EDM directly model the distribution of a target element - it models the distribution of environments predicted to be suitable for occupation, and assumes that mirrors the actual distribution of the element. Even terms like "probability of occurrence", common in discussions of EDM, convey the somewhat mistaken impression that it is the element itself that is being directly monitored and modeled, rather than the actual situation of using a sample of element locations to build an environmental model, and then mapping that model across the study area.

Conceptually, EDM is rooted in the very basic ecological principle that every element is limited in distribution. Elements are not uniform, nor random, in distribution - there are biotic and abiotic factors that constrain where they can and cannot exist in the context of their own biogeographic and evolutionary histories (Pulliam 2000). EDM seeks to describe those limits by correlating element occurrence with environmental factors that directly represent, approximate, or otherwise indicate those limits. It is correlative and descriptive in the sense that occurrence data and environmental data are "allowed" to form the shape of the relationship empirically, as contrasted to more functional or mechanistic approaches that assume a relationship based on physiology or other factors, and then determine how closely the data fit that *a priori* relationship. Therefore, as discussed here, EDM approximates the realized, rather than fundamental, niche of a target element (Pulliam 2000, Guisan and Thuiller 2005, Phillips et al. 2006). The realized niche is discussed in some EDM literature as an element's "multiple response", "ecological profile", "environmental envelope", or similar terms. See Austin (2002) and Guisan and Thuiller (2005) for further discussion on the link between basic ecological theory and EDM.

Most EDM models are static - that is, they estimate and present a constant relationship between an element and its environment, with no temporal component to reflect the more realistic idea that an element's relationship with its environment changes over time (Guisan and Zimmermann 2000). Although temporally dynamic models are possible, they are extremely data-hungry, computationally-intensive, and require an exceptional understanding of nuances in the element-environment relationship. They will not be discussed further in this document. This is the first of many instances in which we raise an EDM issue beyond the scope of this document (and beyond most of its authors, too), and leave readers to pursue it on their own.

EDM involves 3 basic procedures: (1) modeling, (2) mapping, and (3) evaluation (also commonly referred to as "validation") (Rushton et al. 2004; see also Guisan and Zimmermann 2000, Guisan and Thuiller 2005). Depending on the specific modeling approach, these 3 steps can blur together. For example, evaluation may actually occur before mapping (i.e., it is really the distribution *model* that is evaluated, and the final map is the spatial expression of the evaluated model). But in the interest of an ordered discussion, the procedures will be described separately.

2.1 MODELING

First, data on the geographic pattern of occurrence of a given element is summarized in a manner that defines suitable vs. unsuitable environments (or gradations of suitability). In other words, the first step is building a model of distribution based on known occurrences (also commonly referred to as "locality records"). In a broad sense this type of modeling is absolutely nothing new. As a mental exercise it was crucial to the survival of human hunter-gatherers under primitive conditions, although they presumably didn't frame it in terms of "data", "element", and "modeling". One could even argue that some non-

human animals build “models” of the distribution of their forage species using their own repeated “field observations”.

What sets EDM apart from this rather ubiquitous mental modeling is that EDM is pursued in the context of modern science, which means (among other things) that it strives to be transparent, replicable, testable, and objective in results. As with all scientific pursuits, EDM is an application of quantitative reasoning to observational data.

There are two main types of data involved in EDM: occurrences of the target element, and environmental features of the study area. The nature of these data will largely determine the details of the modeling process.

If occurrence data are in the form of qualitative descriptions, modeling may be somewhat coarse. For example, if the only occurrence data available is a published description that reads “species X is found on shaded acidic soils between 3500 and 5000 ft elevation”, and available spatial layers include only landcover, bedrock geology, and elevation, one possible model would define suitable environments as all forest and shrub-dominated patches on granitic bedrock between 3500 and 5000 ft. Unsuitable environments would be all else. It may not be as detailed as one might like, but it is the best model possible given the data at hand. More complete knowledge of the species' habitat preferences or more detailed spatial layers of environmental features could improve the detail and accuracy of the model. Such modeling approaches are described as *deductive* (Corsi et al. 2000), and will be discussed in more detail in Section 4.4.1

More detailed occurrence data can be used in *inductive* modeling approaches. In this case mapped occurrences of the target element, such as Source Features and Element Occurrences maintained in state Natural Heritage program databases, allow statistical modeling of distribution. Each mapped occurrence is plotted on the environmental layers of the study area and attributed with the respective environmental values. Those values are then modeled by a selected statistical function, and the resulting model is mapped across the study area. Statistical models may be enhanced by contrasting the environmental values at points of known occurrence to environmental values at points of suspected absence, or even at randomly available points in the study area. See Section 4.4.2 - 4.4.7.

State Natural Heritage programs and similar entities can efficiently access large amounts of opportunistically-collected occurrence data, usually from multiple observers operating over long time spans, and thus have a great opportunity to use such data in inductive EDM approaches. It is recognized that EDM based on such data is not as powerful as EDM based on data collected specifically for the modeling project itself (Guisan and Zimmermann 2000, Rushton et al. 2004, Reese et al. 2005, Barry and Elith 2006, Elith et al. 2006). Most biologists intuitively understand that a standardized sampling scheme that carefully distributes sampling effort across geographic and environmental gradients would produce the best data for defining the realized niche of a given element in a given study area.

Most biologists also intuitively understand that real limits on time and money mean that waiting until such sampling is implemented would unacceptably delay production of predictive distribution maps for most elements of immediate management concern. For example, Johnson et al. (2004) report a nice model of the distribution of mountain caribou (*Rangifer tarandus caribou*) based on occurrence data collected specifically for the modeling effort. Data collection involved 5 years of helicopter-assisted animal capture followed by repeated telemetry re-locations from fixed-winged aircraft, the total cost of which was likely several orders of magnitude more than what is typically available for the study of most species of conservation concern.

In our experience opportunistically-collected occurrence data, carefully assessed and processed for EDM purposes, can produce accurate and usable models of element distributions. Importantly, this conclusion is increasingly supported by formal studies such as Elith et al. (2006), Phillips et al. (2006), and P. Hernandez (University of Toronto, unpublished data). The point is not to forego standardized data collection in support of EDM, but rather to use existing data to produce predictive distribution maps for use in the short term until more formalized data collection and modeling can be accomplished.

2.2 MAPPING

Just as distribution modeling in the broad, subjective sense is nothing new, neither is distribution modeling in the scientific sense. It has long been a staple of ecology, biogeography, and wildlife science (Rushton et al. 2004). But for most of the history of these endeavors, distribution models have had to remain as verbal descriptions, statistical formulae, or charts that were occasionally mapped over small areas by hand, or more rarely over large areas by the few computer systems and experts capable of such work. More typically, distribution models were applied site-by-site in the field.

However, the recent explosion in desktop computing power, GIS technology, and remotely-sensed environmental layers has greatly enhanced the ability of scientists to map distribution models across real landscapes (Guisan and Zimmermann 2000, Rushton et al. 2004). In other words, distribution models can now be more easily extrapolated spatially, as maps, rather than having to remain abstract as formulae or qualitative text. At the risk of opening a debate on the merits of GIS, we will say that recent advances in GIS technology have not created any really new fields of science, but rather have allowed us to extend and better apply old methods by making them spatially explicit.

Even though distribution models can now be expressed as maps, it is critically important to communicate that such maps are still just models in the sense that they do not show actual, but rather predicted, distribution of a given element. This is easily lost on some end users, because they tend to assume that maps are direct representations of landscapes and thus have little or no error. Most scientific models, hypotheses, and conclusions can be misused by ignoring their inherent uncertainty (which is usually described in excruciating detail and pleading tones by scientists, but is ignored by users anyway), and maps are especially prone to such misuse. A good discussion of mapping uncertainty is presented by Elith et al. (2002).

2.3 MODEL/ MAP EVALUATION

Because of the high potential for mapped models to be misused, it is the responsibility of the modeler to quantify and communicate the uncertainty inherent in any predictive distribution map. This is a critical step in EDM. A distribution model, like any scientific model, should be framed as a testable hypothesis, and until tested its validity remains in question (Verbyla and Litvaitis 1989, Fielding and Bell 1997). In the context of EDM this is usually termed model evaluation (alternate terms in the literature: verification, validation, accuracy assessment), and can be done in a variety of ways that will be discussed in more detail in Section 4.5.2.

Briefly, models can be evaluated in 3 basic ways. First, biologist review and gut-reaction is a potent, and sometimes overlooked, method of model evaluation. The vagaries of occurrence data, mysteries of remotely-sensed environmental data, and oddities of historical biogeography can sometimes combine to produce predictions that are just plain wrong (Williams 1996), and expert review can sometimes detect this better than quantitative analysis. The second, and most prevalent, evaluation technique involves using the model to classify independent occurrence data and calculating various metrics of model

predictive accuracy (Fielding and Bell 1997). Lastly, some argue that the most direct and rigorous way to test a distribution model is via post-modeling field survey: i.e., does the target element occur where the model predicts it should, and does it not occur where the model predicts it should not?

Importantly, practical constraints of data quality, data quantity, and survey resources force some distribution models/ maps to remain un-validated. There are many elements of management concern for which there are only a very few points of known occurrence, very little life history information, and perhaps even few if any experts willing to forward an opinion on model quality. All hard data may be needed to produce the best model possible, and none can be held back for independent validation. Unless validation can be done by field survey, any model in this situation will remain an untested hypothesis and should be explicitly disseminated as such.

2.4 A WORD ON MODEL/ MAP USE

What are the appropriate uses of element distribution models and maps? This is closely tied to the issue of model uncertainty and evaluation. A well-documented model, with high predictive accuracy that has been quantified through rigorous evaluation, can be used for a variety of purposes: guiding field surveys, informing resource management decisions, framing and testing hypotheses on patterns of biological diversity, predicting changes under alternate future scenarios, and others. Un-evaluated models, and evaluated models with low predictive accuracy, can and will be used for the same purposes but their contribution to these pursuits is less; i.e., decisions based on accurate and well-evaluated models are more defensible than those based on less accurate or un-tested models. This again emphasizes the responsibility of the modeler to deliberately deliver models and maps alongside reports that detail their construction, testing, and accuracy (Verbyla and Litvaitis 1989, Fielding and Bell 1997, Loiselle et al. 2003).

Along with overall predictive accuracy it is extremely important to communicate the spatial resolution of element distribution maps. Probably the most common misuses and criticisms of distribution maps grow out of attempts to apply them at much finer spatial scales than they were originally intended. For example, resource managers in western North America commonly complain about the quality of land cover and vertebrate distribution maps produced by the USGS Gap Analysis Program. In almost all cases the complaints come from managers trying to use these maps, which were intended to show regional- and state-wide distributions of flora and fauna, to inform management actions on a hectare scale (M. Jennings, USGS Gap Analysis Program, personal communication).

In a strict sense the coarsest spatial resolution of the occurrence and environmental data used to build a model is also the finest resolution of that model's predictions. If that resolution happens to be 100 ha, as is the resolution of the land cover layer produced by the first Wyoming Gap Analysis (Merrill et al. 1996), then element distribution maps based on that layer are only "good" to that scale. Any activities that require finer-scale information (e.g., where in this 10 ha mineral lease can we put the oil-well pad and still minimize disturbance to sensitive species?) requires on-site observations by field biologists.

In a more practical sense, realizing that even small spatial errors in all of the occurrence data and environmental data used in EDM will interact to produce large, variable, and unpredictable spatial error in the final map (see Barry and Elith 2006), even the best element distribution maps are not appropriate for positioning on-ground actions on a 1 - 100m scale. Said differently, good element distribution maps can identify general landscapes where the target element likely occurs and likely does not occur, but the patterns of occurrence within such landscapes will need to be determined by field observations. The dimensions of the "minimum landscape" are roughly determined by the coarsest resolution of the predictor data used to model the distribution, but are likely some degree coarser than that.

It is not clear how to go about estimating this scale. It is somewhat of an art - efficiently described by Rykiel (1996; as cited in Guisan and Zimmermann 2000) as "discovering the domain over which a validated model may properly be used" - and it will vary by element, study area, and several aspects of data quality (Williams 1996, Guisan and Thuiller 2005). In the end, discovering the scalar limits to EDM predictions may always involve healthy doses of judgment by field biologists familiar with the element and study area. This is very important to communicate to potential map users, and although it causes consternation among GIS specialists and natural resource administrators, it provides some comfort to the muddy-booted among us who worry about job security in an age of computers.

Finally, it is rather common for predictive distribution maps to be misinterpreted and misapplied as habitat quality maps, in the sense of "everywhere mapped as suitable for occupation must be high-quality habitat". Ecologists and natural resource managers too easily conflate the separate concepts of distribution and habitat quality. Species commonly occur in places where they cannot achieve positive rates of reproduction or survival (conversely, dispersal-limited species are commonly absent from places where they could achieve positive rates of survival and reproduction) (Pulliam 2000). Even probability of occurrence, and density, do not consistently predict habitat quality in every case (see Van Horne 1983). If modeled and mapped carefully, with adequate data, a species' distribution might be safely assumed to encompass most high-quality habitat in a study area. But identifying and mapping such habitat requires more work, including considerations of population-level characteristics and vital rates.

3.0 WHY DO ELEMENT DISTRIBUTION MODELING?

EDM can directly contribute to research, management, and conservation, and there is some discussion in the literature as to which modeling approaches are best for each purpose (see Guisan and Zimmermann 2000, Elith and Burgman 2003). As a brief nod to pure research, good distribution models can provide at least suggestive insight into element biology - i.e., knowing which environmental features correlate most strongly with distribution is the first step to finding out why they do so.

Good distribution maps can greatly increase the efficiency of resource management and field research by identifying areas most (and least) likely to support elements of concern. This is particularly important in the applied arena. Knowing where elements of concern likely occur and likely do not occur is a very basic part of natural resource management (Rushton et al. 2004). Development and extraction of natural resources must, by law, proceed in ways that do not unduly affect sensitive elements. When used as planning tools, good distribution maps allow siting of development activities to minimally impact protected biota. Conservationists need to identify not only those environments important to particular elements but also those environments important to multiple elements. Good distribution maps can locate such hotspots efficiently across large landscapes (e.g., Godown and Peterson 2000, Peterson et al. 2002). Public land managers have the privilege of satisfying the demands of everyone, from mustachioed resource profiteers to bring-back-the-mammoth preservationists. Good distribution maps are critical to formulating good compromises (best defined as "plans that leave everyone dissatisfied"). Furthermore, natural resource professionals from across the spectrum face the common problem of having too few resources to survey too large an area for too many elements in too little time. Good distribution maps can clearly help prioritize areas for efficient field inventory.

Throughout most of the 20th century element distributions were typically depicted in 1 of 2 ways: an all-encompassing range map, or a "dot map" showing points of known occurrence. The former is common in field guides, and is usually based on a broad polygon that surrounds the outermost points of known occurrence and occasionally some outlying areas of suspected but unconfirmed occupation. There is usually some degree of subjectivity in plotting a range map. Unoccupied areas within the range

boundaries may be indicated if they are large enough, but more typically the within-range variations in occupation are not shown. Range maps are intended to show possible distribution at coarse spatial scales (regional or continental), and thus are very limited in their ability to inform on-ground actions. Essentially, coarse-scale range maps are exceedingly simple models of field observations.

In contrast, dot maps of points of known occurrence present only hard field observations with no simplification, synthesis, or extrapolation whatsoever. They are similarly limited in their ability to inform on-ground actions, primarily due to the fact that for any given element the distribution of points of known occurrence reflects not just the element's distribution but also the distribution of sampling effort. Said differently, it is difficult to know if the blank areas on a dot map indicate unoccupied, un-sampled, or under-sampled sites (Vaughan and Ormerod 2003). For elements with only a few points of known occurrence, such as many elements of conservation concern, this can be a very significant problem. Also, the past occurrence of an element at a given point does not necessarily mean that it occurs there now, nor that it will occur there again in the future. Careful modeling of an entire set of occurrence points on underlying environmental features can help identify such "false positive" outliers, but simple dot maps leave such decisions up to the user.

EDM is a compromise between the uninterpreted data of dot maps on the one hand and the oversimplified data of range maps on the other. Unlike dot mapping, EDM doesn't end with raw occurrence data. It summarizes those data in an environmental model. And unlike range mapping, EDM doesn't model by a simple encompassing polygon. It relies on quantitative analyses of multiple environmental characteristics at points of known occurrence in a given geographic area. For a species whose entire range is modeled, the result is a prediction of range boundaries as well as predicted variations in occupation within those boundaries - in other words, an estimate of distribution as well as range.

4.0 HOW TO DO ELEMENT DISTRIBUTION MODELING

This is not a complete step-by-step guide to producing element distribution models and maps, but rather a general protocol and a preview of some of the major issues. Although this section is organized as a sequence of steps, some of the steps can be done concurrently rather than sequentially (e.g., there's no reason to wait until the occurrence data is filtered before starting to collect and process environmental data).

4.1 ASKING YOURSELF "WHAT DOES IT ALL MEAN? WHY AM I HERE?"

There is a maxim that most (but certainly not all…) landcover mappers use that applies very well to EDM: how you go about it depends largely on why you are going about it in the first place. In other words, just as you cannot produce a good landcover map without first knowing the use(s) to which the map is going to be put, you cannot produce a good element distribution model/ map without first knowing the use(s) to which it is going to be put. There are a myriad of considerations in this context; some obvious ones, just to get you thinking, are:

> *Are you estimating historic, current, or potential future distribution of an element?* The answer to this question will greatly influence the amount and type of occurrence data used as model input.
>
> *Can you let the modeling algorithm select the best set of predictor variables from an initial set, or should some predictor variables be "forced" into the model?* Sometimes users (usually spelled F-U-N-D-E-R-S) want to know how an element is distributed relative to specific variables - often

variables that are tractable to management, like road density - in which case those variables need to be hard-wired into the analysis.

What informational resolution is expected in the output? Will users be satisfied with predictions of presence vs. absence, or do they want gradations, or probabilities, of occurrence? Similar questions apply to spatial resolution - do users want predictions of presence/ absence on USDA National Forest units across several adjacent states (a scale coarse enough to allow use of many available environmental layers), or do they want predictions of probability of occurrence within individual timber harvest units (a scale so fine that project-specific data layers may first need to be derived from detailed aerial imagery)?

This kind of big picture thinking is critical to making the right decisions in all subsequent steps.

4.2 OCCURRENCE DATA

4.2.1 COLLECTION, QUALITY CONTROL, AND ARCHIVING OF OCCURRENCE DATA

We have nothing to add here beyond reference to the methods, experience, and expertise of state Natural Heritage programs. We include this section merely to point out that before you model an element-environment relationship, you need data on the element and the environment: typically, mapped locations where the element has been observed, and mapped environmental features that will be used to model those occurrences.

The tired phrase "garbage in, garbage out" applies very well to EDM. All the fancy algorithms, statistical gymnastics, and mapping tricks in the world can't help poor input data. In other words, and to state the obvious, the quality and quantity of occurrence data and environmental data will be the primary determinants of the quality and predictive power of the model and resulting map. From an investment standpoint it makes more sense to commit most resources to collecting good field data and tracking down (or developing) good environmental maps, than to tricking-out statistical software and GIS systems. Some modelers say that of every dollar spent on modeling, 90 cents should be spent on collecting good input data; or, similarly, of every hour spent on modeling, 50 minutes should be spent on collecting good input data (there is probably absolutely no data behind these estimates… they just sound good). But you get the idea - quality and quantity of occurrence and environmental data is most of the battle. Yet more comfort to the fish-squeezers, flower-pressers, and transect-readers among us.

It is also worth mentioning that deductive modeling approaches (see Section 4.4.1) can bypass much of the formal collection and processing of occurrence data. Indeed, most of this Section 4.2 doesn't apply to deductive, but rather only to inductive, modeling. But keep in mind that although deductive modeling relies on more qualitative, and sometimes more subjective, information, it still involves occurrence data, environmental data, and a model that relates the 2. It's just that the data are usually stored in the memories of knowledgeable experts, and have been processed and integrated into element-environment model based on their own experiences. And this data still needs to be collected, documented, and archived in order for resulting models to be transparent and replicable. Expert opinion is neither fast nor easy to acquire and organize. In the end, deductive modelers may spend as much time and effort constructing and populating databases with expert input (see Southwest ReGAP 2005 for a description of such a database) as inductive modelers spend in collecting and processing mapped occurrences.

As discussed previously, opportunistically-collected occurrence data from multiple sources are not as powerful as EDM input as are standardized occurrence data collected specifically for EDM. In the best of

all possible worlds there would be enough time and resources to produce standardized modeling data sets for all elements of concern, but this is not that world.

4.2.2 FILTERING OF OCCURRENCE DATA

Anyone who has managed large sets of biological occurrence records clearly understands that such records require processing before they are used as input to EDM. This processing can be thought of as filtering - weeding out records that, for one reason or another, would end up confusing or biasing the model rather than contributing to it. The big picture thinking in Section 4.1 helps a lot with asking the right filtering questions and making the right filtering decisions.

The below list of filters assumes that occurrence data has been opportunistically-collected by variety of sources and has been previously quality-checked for gross errors such as duplicate records, mis-mapped records, and so forth.

Negative data: If a dataset contains negative data (i.e., absent points; records for locations where an element was sought, but not found), the negative records need to be either removed or somehow flagged so they are not inadvertently included as positive, documented occurrences of the target element. This seems obvious, but data contributors sometimes fail to notify recipients that the contributed dataset is a mixture of positive and negative records. Note that some modeling approaches require negative records (see discussion on absence data, Section 4.2.4), in which case such records could simply be moved to another dataset that enters the modeling process at a different step.

Element identity: EDM strives to produce a spatially-explicit extrapolation of the environments selected by a given element, as modeled by a set of known locations of that element. Mistakenly including locations of other elements in the modeling dataset will degrade the selection signal of the target element. Thus, for those elements that could be easily confused with others (e.g., *Lynx canadensis* vs. *Lynx rufus*), it is a good idea to evaluate the likelihood that each occurrence record represents an actual observation of the target element. Such evaluation can be assisted by identification fields in the contributed dataset, if such fields exist. Records with "unknown", "questionable", "unlikely", or similar entries may, on the opinion of the modeler, need to be removed from the occurrence data set. Evaluating the source of a record may also be helpful - e.g., records from trained field biologists may be retained, whereas those from laypersons or unknown sources may need to be rejected. Obvious outliers in time (e.g., a bat flying around North Dakota in January) or space (e.g., a cactus blooming at 13,000 feet) might get removed as misidentifications, too, although one should be very confident that such records are misidentifications or risk removing important target element locations from the model.

As with almost all filters, most of the decisions in this step will depend on the judgment of the modeler. Is the target element easily misidentified, and if so does its appearance or life history suggest efficient ways of screening out misidentified records?

Mapping precision: This will come as no great revelation to personnel in state Natural Heritage programs: there is always some degree of spatial error inherent in assigning map coordinates to observational data. As stated previously, a distribution map is a spatially-explicit extrapolation of the environments selected by a given element, as modeled by a set of known locations of that element. Precisely-mapped locations may represent environmental choices more accurately than coarsely-mapped locations, and therefore it makes sense to minimize the use of coarsely-mapped locations as much as possible. If each record is labeled with an estimate of mapping precision,

then this can be a relatively straightforward process. It is more difficult when working with contributed datasets that lack precision fields.

What constitutes acceptable precision has a lot to do with the target element, the precision of the predictor environmental data, and the model intent. For example, if the goal is a multi-state model of a wide-ranging habitat generalist based on 1-km climate data, then occurrence records with an uncertainty radius $\leq$500m might be acceptable input. Contrast this to a county-wide model of a rare plant, specialized to very discrete rock outcrops, based on 10-m geology and soils layers. Again, the judgment of the modeler will play a key role in determining acceptable mapping precision.

Historical records: Element distributions are dynamic, some very much so. It is therefore crucial to consider the year-of-observation of each occurrence in the context of the overall goal of the modeling project. If the goal is to estimate the current distribution of an element that has undergone a recent range contraction, then some cutoff date will have to be determined such that records of observations made before the cutoff are not used as model input. Deciding on a cutoff, of course, involves more case-by-case judgment. As with the mapping precision filter, this is more straightforward when records are attributed with dates, and more difficult in the surprisingly many cases when even year-of-observation is not available.

Season of occurrence: This one seems obvious, too, but to avoid having the model recipient point out that your distribution maps of a migratory taxon seems to be a rather liberal union of breeding, migratory, and winter habitat, you might want to double-check that models for migratory taxa are based only on observation records from the season within which the model is supposed to apply.

Extra-limital records: Due to the mobile nature of many organisms, especially animals and in particular migratory birds and wide-ranging mammals, it is sometimes necessary to eliminate extra-limital records (i.e., individuals that have dispersed far outside the areas of anticipated occurrence). Such records can bias models and lead to inappropriately broad distribution patterns. However, it is important that, although such records not be used in current models, they be retained for future consideration, because today's accidental occurrence could be the early stages of a later shift in range.

Keep in mind that there are cases when it is better to retain "poor" records than to eliminate them. For example, in one case a rigorous application of the mapping precision filter would have reduced our sample size of occurrence points from several dozen records to almost none. Given the particulars of the modeling project and needs of the requester, we decided that a model based on many poorly-mapped points was better than one based on only two or three well-mapped points. In another case, applying a strict element identity filter would have eliminated all records from the southern half of an element's range in our study area (a five state region, in this case). Again, after considering the options, we decided to retain those southern records because to eliminate them would have produced a model with an unacceptable bias towards northern environments. Said differently, we decided that the error produced by having a few non-target element observations in the input data would be less than the error produced by ignoring a substantial portion of the target element's range.

The overall goal of filtering the occurrence data is to come up with the best possible modeling input, and evaluating such trade offs is an unavoidable part of the process.

4.2.3 DEALING WITH UNEVEN SAMPLING EFFORT AND OCCURRENCE CLUSTERS

Ideally, EDM is based on points of known occurrence from throughout the entire range of the target element in the study area. Furthermore, to best reflect the element's environmental preferences, such points should result from even application of sampling effort throughout the study area. Needless to say, these conditions are rarely if ever met and almost all distribution maps are based on opportunistically-collected occurrence data that include an unknown degree of sampling bias; i.e., some of the study area has been sampled much more intensively than the rest, with many sites having received little or no sampling for the target element. If not accounted for, the clustering of points in heavily-sampled environments can bias the model/ map towards those environments (Vaughan and Ormerod 2003). One well-known example of how opportunistic data can be spatially-biased is the strong association of observation records with roads. Many road networks are closely traced by plots of observation records, with commonly-used roads showing up especially well. This pattern is less indicative of elements preferring to reside near roads than it is of biologists sampling more extensively along roads than in remote areas.

Therefore, once the occurrence dataset is filtered for negative records, element identity, mapping precision, and so forth, it's a good idea to remove more records in a manner that reduces this clustering bias. Unfortunately there are no general recommendations on how to do this in a practical, effective way. It is not an easy issue to deal with. It has been framed as a problem of spatial autocorrelation (e.g., Smith 1994, Augustin et al. 1996, Guisan and Zimmermann 2000, Rushton et al. 2004), but there is some danger in trying to remove spatial autocorrelation in an occurrence dataset being used for EDM. All landscapes are heterogenous, and all elements will naturally cluster (and thus show spatial autocorrelation) in patches to which they are best adapted. Indeed, that is exactly the pattern we are trying to uncover in EDM! The problem, as stated above, is that some of the spatial autocorrelation is not due to element environmental selection, but rather to uneven distribution of sampling effort. And it is difficult to tease the two apart. Spatial autocorrelation analysis will probably eventually produce good techniques in this respect, but as it stands now there is no clear solution.

We have dealt with the problem in a relatively straightforward manner that is far from perfect, but at least addresses some of the issue. We eliminate post-filtering occurrence points such that no two points are within a certain distance of each other. The separation distance differs by taxon, and roughly reflects the relative mobility of that taxon (e.g., it is larger for birds and wide-ranging mammals, smaller for amphibians and plants). Commonly there are multiple records that are mapped at the exactly same point, and our removal procedure reduces such multi-records to a single record per point location. Also, prior to removing points from a cluster, our removal procedure evaluates the mapping precision of each record to ensure that records with the finest mapping precisions are preferentially retained and those with the coarsest mapping precisions are preferentially removed. In this manner, we at least reduce those clusters of occurrences that arise from some occupied sites being repeatedly visited and reported (e.g., the bald eagle [*Haliaeetus leucocephalus*] nest that all the local birders know about, or the burrowing owl [*Athene cunnicularia*] colony next to a major road intersection) from multiple points to single points. In a modeling context, this gives those environments the same weight as environments represented by fewer observations.

Ultimately, the state Natural Heritage program methodology of building biologically-relevant Element Occurrences (capitals deliberately used to denote the formal EO construct) may have a lot to contribute here. The grouping of individual observation points into EO's, guided by the details of element mobility and life history, goes a long way towards teasing out clustering due to uneven distribution of sampling effort and leaving behind clustering due to environmental selection. For elements whose individual records of occurrence have been summarized into EO's across the study area, using a single point per EO as modeling input may be the best way to go.

Given the advent of the Biotics methodology that allows tracking of EO's as precisely delineated polygons, it may be more powerful to base EDM on a synthesis of environmental attributes across polygons rather than at simple point locations. This is discussed in more detail in Section 5.0 (*Using linear and polygonal analyses).*

A final data processing consideration: once a set of opportunistically-collected occurrence data is filtered appropriately and subsampled to address clustered sampling effort, there are usually only relatively few records left for model input. But in some situations (likely involving common elements and large study areas) there may still be a lot of occurrence data left, and the modeler may want to consider additional subsampling by geographic strata and major environmental gradients (e.g., elevation zones, precipitation gradients) to better approximate a data set generated by a more standardized sampling scheme (Guisan and Zimmermann 2000). To date we have not pursued this approach, nor have we fully evaluated its potential advantages or disadvantages. An obvious question is at what point do the advantages of a more standardized coverage of occurrence points outweigh the disadvantages of a reduced sample size? This additional subsampling is mentioned here only as a potentially useful technique, the merits of which have not yet been assessed (but see Reese et al. 2005).

4.2.4 ABSENCE DATA

A very important consideration is that some statistical EDM techniques, such as classification and regression tree [CART] analysis and logistic regression, require not only data points where the target element is known to occur ("present points"), but also data points where the target element does not occur (negative data, or "absent points"). And one quickly gets into some rugged philosophical terrain when contemplating absent points. Whereas you can prove presence at a given point by observing an element there (but just because you saw it there once doesn't mean it has always been there, or will ever be there again…), you can never really prove absence. Just because you haven't observed an element at a point yet, it could still show up there tomorrow.

Conceptually, mapping species absences quickly leads to the issue of "detectability" of species under field conditions. Statistical models of species detectability are getting a lot of scientific ink these days (see MacKenzie et al. 2006, and papers accompanying Vojta 2005), with some of it directed at least tangentially to EDM. Argaez et al. (2005) and Gelfand et al. (2006) show how formal considerations of detectability can produce confidence maps to accompany predictive distribution maps. These discussions appear to be mostly academic at this point, and have not yet produced any practical tools for the applied EDM modeler.

In a practical sense the issue of absent points boils down to a question of sampling effort. How much sampling effort, of what type and across what time span, is necessary at a given point before you are willing to make the call that the target element likely does not occur there? Some taxa have published guidelines that can help in this regard. For example, the USDI Fish and Wildlife Service (USFWS) has issued trapping protocols for field inventories of Preble's meadow jumping mouse (*Zapus hudsonius preblei*). If the protocol is applied for 750 trap-nights in a given location without capturing an individual, then USFWS is willing to consider the site "cleared" for the taxon. The USFWS has published similar guidelines for field inventories of mountain plover (*Charadrius montanus*) and several other taxa recently under consideration for Endangered Species Act listing.

Without such published guidance, the definition of adequate sampling effort is left up to the modeler. Fertig et al. (2003a, 2003b) derived a clever way to derive absent points for several dozen rare plants simultaneously. The Rocky Mountain Herbarium (RMH; University of Wyoming) archives almost a million specimens of plants collected throughout the Rocky Mountains, with each specimen labeled by

collection location (to the highest spatial precision possible, given the collector's report). Fertig et al. (2003a, 2003b) first reduced the complete set of RMH records to just those within their study area, then tallied the number of collection sites in each land survey section (1 mi x 1 mi) in that area. They reasoned that if a given plant species occurred in a particular section, then it should be represented in the RMH dataset for that section if enough field sampling had taken place there. They set their definition for "enough" at 20 collections; i.e., if >20 collections were made in a given section, and the target plant species was not collected there, then the centroid of that section was designated as an absent point for that species.

Some botanists reading this are already groaning, and most zoologists are rolling their eyes. But keep in mind that no one claims this to be a perfect system, and the pursuit of a perfect set of absent points may be wasted effort. The idea is to get a set of points that are as "absent as possible", under the constraints of time and money, to enter into the model. Given the philosophical strangeness of absence data and the practical problems in generating it, it is good to get in the habit of thinking about and discussing absent points as "points of suspected absence". The Fertig and Thurston (2003) approach is presented here as one way to go about it. Their reasoning may spark ideas for deriving points of suspected absence for other elements in other situations.

Of course, there's always the option of forgetting absence altogether and generating a set of randomly-located "available points" (also called pseudo-absence points) across the study area with which to contrast present points. Generating pseudo-absence data is topic that has recently gained a considerable amount of attention in the literature (e.g., Zaniewki et al. 2002, Anderson 2003, Engler et al. 2004; see also MacKenzie et al. 2006, and papers accompanying Vojta 2005). Comparing presence vs. available is not as powerful as comparing presence vs. suspected absence, but in many cases there is no practical way to generate points of suspected absence. We have used the presence vs. available approach for modeling distributions of vertebrates. In some cases we have first generated randomly-placed available points, then added the few points of suspected absence that were available, under the general thinking that the set should be as absent-as-possible.

When generating a set of absent or available points it is important that such points cover the entire study area. After all, the goal is to model distribution across the entire area, so the statistical contrast should cover the area, too. When we were producing a statewide model of pygmy rabbit (*Brachylagus idahoenesis*) distribution we generated a set of available points that covered the whole state. We even went so far as to make sure each elevation band in the state was proportionally represented in the available set. But pygmy rabbits occur only at low elevations in the southwestern quarter of Wyoming, and it was suggested that we restrict our distribution of available points only to that area. We reasoned that would be incorrect because our goal was to model distribution statewide. If we were going to model distribution just within the low elevations of southwest Wyoming, we would restrict our point sets to just that study area. Basically, such decisions should be made in light of core project goals, as suggested in Section 4.1.

It is also important to apply the reasoning in Sections 4.2.1, 4.2.2, 4.2.3, and 4.2.5 to absent points and pseudo-absence points. These point sets should be carefully generated, filtered for incorrect and vague records, separated by some minimum distance to minimize biasing a model unnecessarily against particular environments, and split into modeling and evaluation sets via geographically stratified random sampling.

4.2.5 SPLITTING OCCURRENCE DATA INTO MODELING AND EVALUATION SETS

At this point the modeler should decide whether there are enough occurrence data (including absent points, if necessary) left to hold some back as an independent dataset with which to quantitatively evaluate the model, or if there are so few occurrence points that they all have to be used for modeling (and the model will remain, and have to be delivered as, an untested hypothesis). This is another judgment call. We've produced some good models with as few as 10 occurrence points, and we've produced some mediocre models with many times more. It is a case-by-case, or more precisely element-by-element, decision. In the opinion of a knowledgeable biologist, does the post-filtering dataset reasonably approximate the element's range and distribution? Is the element so rare that acquiring more occurrence points is unlikely, or impractical, in the short run? Is the element a habitat specialist, or generalist, or is that even known? Is the end-user going to be comfortable with just letting 'er rip on the full set of points and pushing evaluation to a later phase of the project, or do they require formal evaluation (see Section 4.1 again)?

If the modeler decides to split the occurrence data for quantitative evaluation, the next step is to determine how many data points to place in the evaluation set. As a general rule we shoot for about 75% of the occurrence data to remain in the modeling set (also known as the training set or calibration set) and 25% to be selected for the evaluation set (sometimes also called the validation or testing set). Of course, this rule can be bent according to circumstances and the judgment of the modeler. There really is no standard here, although one should probably consider whether an evaluation dataset of <10 points can adequately test a model.

When selecting points to go into an evaluation dataset it is important to balance random selection with spatial extent. In other words, it is a good idea to perform stratified random sampling, whereby points are selected randomly within geographic strata. Pure random sampling can select data points from a relatively small portion of the element's range in the study area, especially if there are few points in the evaluation set. Such a set would not be a good test of the model's predictive power across the whole study area. It is relatively easy within a GIS to place a grid over all of the mapped occurrence points, and then randomly select points from each grid cell to place into the evaluation set. In this manner the spatial extent of the evaluation set mirrors that of the modeling set. Indeed, there is some suggestion in the literature (e.g., Elith et al. 2006) that modelers may consider structuring their evaluation datasets even more strictly, with stratification across predictor variables and other gradients besides simple geographic space.

At this point it is important to mention two variations on model evaluation that have implications for the splitting of occurrence data into modeling and evaluation sets. First, some researchers (e.g., Fielding and Bell 1997) advocate a modeling approach whereby (1) occurrence data is split into modeling and evaluation sets, (2) a model is built on the modeling set, (3) quantitative evaluation proceeds using the evaluation set, and (4) a final model is built using all of the occurrence data pooled back together. This approach has some intuitive appeal, in that the final model is based on the full predictive power of all of the occurrence data. However, it is clear that the quantitative measures of model accuracy will not directly relate to the final model but rather to the "pre-model" built only on the modeling set. Because these measures are estimates in the first place, divorcing them further from the final model gets one into the error propagation problems of "estimates of estimates"; i.e., how well do they really estimate the accuracy of the final model?

Second, there is increasing interest in evaluating EDM output by statistical re-sampling (think back to those jackknifing and bootstrapping lectures; Verbyla and Litvaitis 1989, Fielding and Bell 1997; Rushton et al. 2004). As with the pre-model technique outlined above, re-sampling is appealing because the final model is based on almost all of the occurrence data rather than a subset. This is discussed in more detail in Section 4.5.2, but suffice to say at this point that although EDM evaluation by re-sampling

has a lot of potential it does not appear to have yet been fully researched and, most importantly, its implementation is not currently supported by easily-used software routines.

4.3 ENVIRONMENTAL DATA

4.3.1 COLLECTION OF ENVIRONMENTAL DATA

To this point most of the discussion has been about occurrence data - the dependent variable in EDM. The other side of the coin is environmental data - the independent, or predictor, variables, in EDM. These days there are a lot of environmental layers available in digital form, and each has its particular advantages and disadvantages, strengths and weaknesses.

Recent advances in GIS and remote sensing technology, coupled with government, university, and private support, has produced a solid industry generating increasingly detailed spatial layers of environmental features. At the risk of glossing over some important issues, we'll admit that we have essentially just followed the lead of this industry. For our current EDM projects we use spatial layers that are already tested and available, complain a lot about those that aren't, and develop very few layers on our own. This is due largely to the fact that developing accurate, high-resolution data layers from original remote sensing information for areas as large as states (or even counties) is a huge undertaking requiring sizeable amounts of funding, computing power, and GIS expertise, not to mention field validation. We are often able to adjust existing layers to better meet our EDM needs, but derivation of entirely new data sets is largely out of reach. Those that we have developed tend to be rather simple combinations of existing layers (e.g., smashing together layers of roads, mines, wells, railroads, powerlines, and pipelines to produce a "disturbance index" layer).

There are several environmental layers that we wish we had. For vertebrate EDM we could really use a consistent, state-wide coverage of simple vegetation structure, like canopy density. For plant EDM a more precise map of soil types would really help. Guisan and Zimmermann (2000) suggested that improving accuracy and resolution of layers of qualitative features like soil type and land use will greatly improve EDM predictive accuracy in general. Clearly there are ways to work with remote sensing and mapping specialists to improve EDM work in the future, and this should be a primary focus of long-term EDM development - refer back to the "garbage in, garbage out" discussion in Section 4.2.1.

Environmental layers used as predictors in EDM must completely cover the study area (and ideally do so because a consistent mapping methodology was applied across the study area) because the goal is not just to attribute points with particular environmental values but then also to map the resulting model across the study area. Mid-way through an EDM project it is common for people to suggest field visits to all occurrence and validation points in order to measure fine-scale environmental variables in the field - maybe soil depth, or some detailed component of vegetation structure like understory height - so those variables can be included as predictors in the model. But whereas they might contribute to the model, and indeed might be wildly significant predictors of presence of the target element, there will be no way to map that model unless there are complete maps of soil depth or understory height, derived from methods identical to those used in the field, that cover the whole study area.

4.3.2 SELECTING ENVIRONMENTAL PREDICTORS FOR MODELING

When compiling a library of environmental layers for EDM it makes sense to cast a fairly wide net to gather all potentially relevant data. However, selecting a subset from this library to use for a particular

model requires careful reasoning. At the risk of oversimplifying, there are essentially two schools of thought on how to select the right predictors: model calibration (after Guisan and Zimmermann 2000) and reasoned selection.

Model calibration (aka model fitting; aka model tuning; aka statistical fishing) involves any number of iterative routines that build and compare many possible models, each using different subsets of predictor variables, to identify the one or few best models. Some routines don't go all the way to building and comparing final models, but rather do more general statistical comparisons of variable sets - in a sense they model potential models. Some calibration routines are quite sophisticated (e.g., Cumming 2000, D'heygere et al. 2003), and as a result require a lot of time, expertise, and computer power.

In contrast, reasoned selection proceeds from biological knowledge of the target element and study area. Essentially, biologists familiar with the target element and study area identify predictor variables that are known or suspected to be either relevant to element distribution or explain biologically-important variation in the study area.

As with all process dichotomies, the most effective approach probably blends these two extremes. It is appropriate to begin the process with reasoned selection, if for no other reason than to reduce the bewildering array of all possible predictor variables (most biologists can probably quickly identify >10 potentially relevant variables for any element and study area) to a manageable number. Obviously, if there are environmental variables known or suspected to drive the distribution of the target element, then layers of those variables (or, if those aren't available, then layers that approximate or indicate them) are strong candidates for the modeling set. But if those variables don't vary much across the study area, they won't contribute much predictive power to a model. So it is a good idea to preview each of the potential predictor variables to see how they map across the study area, and consider using those that show strong spatial patterns. Finally, sometimes the end-user of the model wants to see particular variables included, so they need to be hard-wired into the analysis (see Section 4.1 again).

In lieu of some of the arcane calibration techniques in the primary literature, some simple variable reduction procedures like principal components analysis (PCA) can then help sort through the predictors that make the first reasoned cut (Rushton et al. 2004). A PCA of all variable values in the study area will indicate which variables explain the most variation (i.e., which variables load highest on the first few principal components) and which variables are inter-correlated (i.e., which variables load highly on the same principal components). After first selecting those variables that "need" to be in the final set (as determined by element biology or end user needs), the loading matrix can help guide selection of additional variables that explain the remaining variation in the study area.

There are variations on this procedure. We've run 2-stage PCA's on climate data to help select variables for regional-scale projects. A preliminary PCA on a set of a dozen temperature variables identified the best 4 or 5; similar preliminary PCA's selected the best sets of precipitation, humidity, and radiation variables. A final PCA on the variables selected by the preliminary analyses then identified the final set of climatic variables to enter into the models.

Increasingly, information theoretic approaches such as Aikaike's information criterion (AIC) are being used to compare multiple candidate models (e.g., Stephens et al. 2005, Greaves et al. 2006). AIC is a rather attractive tool in this regard because it balances model fit with parsimony - models with many predictor variables are penalized relative to those with fewer. It also requires the modeler to identify a relatively small number of predictor variables to start with - i.e., it proceeds from a foundation of reasoned selection (Rushton et al. 2004). AIC and other techniques that rank multiple candidate models are especially attractive when modeling the distribution of poorly-studied elements in poorly-studied

areas. These conditions justify a deeper exploration of the predictive power of different variables and variable combinations.

In summary, models based on carefully selected predictor variables are more defensible, and usually predict better, than models and maps based on pure statistical fishing (McCarthy and Elith 2002, Rushton et al. 2004). Some modeling algorithms will gladly overfit occurrence data to environmental data - the more environmental layers added, the more overfitting they'll do (it's analogous to increasing r^2 every time another independent variable is added to a multiple linear regression). Furthermore, models based on straightforward variable selection are more easily explained, more easily understood, and thus more likely to be used. So there are good reasons to keep the set of predictor variables fairly tight.

Of course, independent variables don't have to be directly related to environmental features. Principal components themselves, or other mathematical combinations of individual variables ("orthogonalized components derived from environmental variables through multivariate analyses"; Guisan and Zimmermann 2000), can be used as predictors. This approach should probably be used with caution, however. Although synthetic variables can (often marginally) increase the predictive power of a model, they are difficult to interpret and explain to model users. They have an aura of mathematical black art. For example, most natural resource managers can understand a model and map that show a species' distribution as a function of landcover type, elevation, and total annual precipitation. But it's much harder to wrap your head around a model and map that show a species' distribution limited by values of <2.10 on principal component #1 and >-0.05 on principal component #2. The intuitive feel for landcover type, elevation, and precipitation is lost when they are converted to synthetic variables.

By the way, the same kind of thinking can be applied to some of the more creative environmental variables generated by remote sensors and landscape ecologists. In short, if you're going to use "double-log fractal index" or "area weighted mean class edge contrast" or similar variables in a model, you should be prepared to explain what they are, why they are important, and how a manager can work with them.

4.4 MODELING ALGORITHMS

At last, there's the modeling. It's worth reflecting why it's taken so long to get here: EDM is one of those "pyramid" pursuits that requires a strong foundation of careful preparation and planning, to the extent that the actual modeling is a relatively small part of the project. That said, there is still a lot of thought that must go into selection of the appropriate algorithm and, moreover, the type of algorithm used often influences how one must consider all the preceding information.

Below we've organized several modeling techniques into groups. We do not claim that this is the best organization (see Guisan and Zimmermann 2000, Elith and Burgman 2003, Seguardo and Araujo 2004 for other schemes), nor is it necessarily complete. Please refer to recent publications such as Elith et al. (2006) for more exhaustive lists of modeling algorithms in use today. Below we discuss deductive modeling and four relatively robust inductive techniques that we want to regularly apply in EDM projects: DOMAIN, CART, multiple logistic regression, and Maximum Entropy. We also briefly mention Bayesian modeling approaches, because of the increasing attention they are receiving over traditional methods.

4.4.1 DEDUCTIVE MODELING

Deductive modeling, in which element occurrences and element-environment relationships form out of correlative ecological studies or the field observations and experiences of qualified experts (rather than

mapped occurrence data), is a valuable EDM technique under certain circumstances. Such models can be produced and mapped relatively quickly (in part because they bypass much of the formal data collection and processing outlined above), making them attractive for conservation planning projects that involve large areas and many elements (e.g., USGS Gap Analysis [Scott et al. 1993, Merrill et al. 1996, Davis et al. 1998]; see also Kautz and Cox 2001). Also, in situations where there is very little hard occurrence data or the collection of such data would be too expensive, expert opinion models may be the only EDM option (Argaez et al. 2005). But recall the discussion from Section 4.2.1 - in order for deductive models to be transparent and replicable, their constituent expert input must be documented in an organized fashion. This may be easy for poorly-studied species, but may require much time and effort for better-known taxa.

Deductive models are sometimes looked down upon as too subjective and not replicable or quantitative enough for real science. Most of this sentiment seems to echo from towers of somewhat ivory tone. The real world of applied natural resource science usually involves practical constraints of time, money, data, manpower, and expertise, which can preclude the use of the more demanding inductive modeling techniques. However, there is an emerging sense that inductive models may be better than deductive models in most cases (Pearce et al. 2001, Seoane et al. 2005), and thus inductive approaches should probably be used whenever possible.

(As a parting shot at the ivory tower, it is interesting that the frontier of quantitative modeling is in artificial intelligence - genetic algorithms, neural networks, and other processes that explicitly try to capture the process of human reasoning. From a practical standpoint, we all know several intelligent, genetically-based neural networks that are pre-programmed with lots of biological expertise and are usually quite willing to share it… we might as well use them in EDM when we have the need. It has been pointed out that their user-interfaces can be quirky, though.)

Deductive models typically synthesize knowledge of element distribution gathered via interviews, literature survey, or other techniques. The synthesis is usually operationalized into discrete relationships between the target element and the environment: e.g., species Y occupies land cover types 1, 2, and 4, and does not occupy land cover types 3, 5, or 6. Such information can also be operationalized in more continuous relationships, often derived from more rigorous field observations: e.g., the probability of occurrence of species Y increases with elevation from 3000 to 6000 feet, then remains steady until 8000 feet at which point it drops rapidly to 0. The expert or modeler may end up literally drawing the relationship themselves, as in the traditional Habitat Suitability Index system (USDI Fish and Wildlife Service 1981; see also Short et al. 1996).

A theme that will re-occur in the discussion of all modeling algorithms, and that is extremely relevant to model evaluation, is the selection of a model threshold value that separates suitable from unsuitable environments. To briefly preview model evaluation, most (but not all…) basic evaluation techniques require that model output is converted into a binary form: suitable vs. unsuitable environments (Guisan and Zimmermann 2000). This forms the template over which occurrence points in the evaluation data set are laid and classified, which in turn forms the basis for estimates of model accuracy. The modeler therefore must select the threshold value of the predictor variable(s) that divide suitable from unsuitable environments. This is most easily envisioned for a single-variable model - e.g., for a given boreo-alpine vertebrate, what elevation best separates unoccupied lowlands from occupied highlands? Clearly, multivariate thresholds are required for most EDM projects.

Deductive models are no different from other model types in this regard. They require threshold selection for basic evaluation measures. In this context it is important to remember that EDM involves modeling in a somewhat pure sense, with no statistical inference or significance values involved, so the modeler has quite a bit of control over the final form of the model. Careful attention should be paid to making

threshold decisions and establishing other calculation rules in defensible manners. Threshold selection can be somewhat of a hidden step in deductive modeling, because the approach tends to blend model formation and threshold selection. That is, when an expert forwards an opinion on a significant relationship, he/she usually also implies a threshold value.

4.4.2 INDUCTIVE MODELING - BIOPHYSICAL ENVELOPES

Inductive, statistical models differ from deductive models in that the EDM process is more objective and data-driven. The output is determined by how the occurrence data plots onto the environmental variables, and how that pattern is interpreted by the selected statistical function (with all of its attendant assumptions). There are many benefits to inductive modeling. For instance, inductive models don't necessarily require expert knowledge of the species biology (which may be biased or lacking; Seoane et al. 2005), they use occurrence data that is increasingly available (e.g., computerized and georeferenced collections), they are scalable, and they are easily updated when new point data becomes available. However, they also require more computational power, employ more sophisticated statistical algorithms, and require careful attention to model assumptions and evaluation.

There are a myriad of statistical techniques used in inductive EDM. One class can be termed "biophysical envelopes". Envelope techniques essentially bound a set of occurrence points in the multivariate space defined by the environmental predictor variables (Guisan and Zimmermann 2000, Farber and Kadmon 2003). Portions of the study area that fall within this envelope are interpreted as suitable for the target element, and portions that fall outside are interpreted as unsuitable (recall the discussion of threshold selection, above). There are several envelope modeling algorithms available as "canned" packages - BIOCLIM was one of the first, and is probably the best known. Some are already integrated, or are built to be easily integrated, into a standard GIS for relative ease of use.

Most of these canned packages are accessed through dialog boxes, the first of which usually prompt the modeler to identify the files containing the occurrence data and also the environmental data he/ she wishes to use as predictors (naturally, the data has to be in particular formats, file types, geographic projections, etc.). The modeler will also be prompted to select certain calculation rules and threshold values. For example, for each point in the study area, the DOMAIN algorithm calculates a similarity index based on how closely the environmental values at that point match the environmental values at points of known occurrence (Carpenter et al. 1993). It is up to the modeler to decide how close is close enough; i.e., which similarity value is the threshold, such that points with values above the threshold are within the suitable envelope and those with values below the threshold are outside of the suitable envelope. For past DOMAIN models we've routinely set the threshold at the similarity value that places 95% of the points of known occurrence in the modeling data set within the suitable envelope (or, in other words, the similarity value that excludes 5% of the most dissimilar points of known occurrence). We are currently moving toward a more quantitative determination of optimal thresholds via certain evaluation statistics (see Section 4.5.2).

After identifying file locations, thresholds, and other rules, the user launches the program which runs flawlessly, never crashes, and always produces exact results… … …

Comparisons of several envelope modeling routines suggest that DOMAIN produces accurate models under a variety of circumstances and can be rather robust to low sample sizes of occurrence data (P. Hernandez, University of Toronto, unpublished data; see also Elith et al. 2006). We have used DOMAIN to successfully model the distribution of rare plants and vertebrates in Wyoming and the surrounding region, and plan to continue to use it as our primary envelope modeling technique. One big advantage to

DOMAIN is that it does not require absence or pseudo-absence data (Seguardo and Araujo 2004). The modeler need only bring presence data to the DOMAIN algorithm to get a predictive map.

4.4.3 INDUCTIVE MODELING - MULTIPLE LOGISTIC REGRESSION

Most natural resource scientists are familiar with modeling the relationship between species occurrence and environmental gradients via simple and multiple linear regression. Most have also been exposed to simple and multiple logistic regression (Hosmer and Lemeshow 1989). Logistic regression has the advantage of modeling a binary response variable in terms of probability of achieving 1 of the 2 states given certain values of the predictor variables. Or, in the context of EDM, calculating the probability of presence of the target element given the values of environmental predictor variables (see Augustin et al. 1996, Beauvais and Smith 2003, Pearce and Ferrier 2000a).

Until very recently there were no canned packages that performed EDM with multiple logistic regression, which meant that data had to be moved by brute force between a GIS and a statistical software package. The modeler first attributed occurrence points with environmental values in GIS, then exported that data to the statistical software to run the model, then modified and combined the pertinent environmental layers according to the model coefficients to express the model as a map. In May 2004 the StatMod Zone extension to ArcView (http://www.gis.usu.edu/~chrisg/avext/) was made available, which automates much of this work if the modeler has the appropriate statistical software and other extensions. The IDRISI GIS has some limited logistic regression capability as well.

Recall the previous discussions of user-selected thresholds and calculation rules. The analogous problem in logistic regression modeling is the selection of a probability threshold such that sites with probability values above the threshold are denoted as suitable for the target element, and those with lower probabilities are denoted as unsuitable. There are some guidelines here - for example, Fielding and Haworth (1995) selected the threshold as the mid-point between the average probabilities of all present points and all absent points in the modeling data set. In the end, threshold selection is under the modeler's control and may require case-by-case considerations.

Finally, note that logistic regression requires absence (or pseudo-absence) data in addition to presence data.

4.4.4 INDUCTIVE MODELING - CLASSIFICATION AND REGRESSION TREE

CART modeling (Breiman et al. 1984) is a discriminant process wherein a set of points of known presence and a set of points of suspected absence (yep, you need absence or pseudo-absence data for this one, too) for an element are successively split from one another. The environmental variable, and value of that variable, that best divides all points into a "mostly present" subset and a "mostly absent" subset is chosen first. Those 2 subsets are then each split again into "mostly present" and "mostly absent" subsets using the variables that best do that for each set, and so on until the final subsets contain only points of known presence or only points of suspected absence. The final product, then, is a dichotomous tree that shows a series of cut-points on environmental variables that lead to suitable (i.e., those pathways that lead to subsets of only points of known presence) and unsuitable environments (i.e., those pathways that lead to subsets of only points of suspected absence).

An important output of CART modeling is not only a dichotomous tree model that can be mapped, but also an indication of the importance of the predictor variables in the final model (Prasad et al. 2006). The first variable in the output tree (the one that best divides the full point set into mostly-present and mostly-

absent subsets) can be interpreted as the most important variable in determining species distribution. The variables that form the next splits in the tree are next most important, and so forth. This feature makes CART useful as a model calibration or "variable exploration" tool.

Also, in an important contrast to biophysical envelopes and logistic regression, CART can identify multiple suitable environments in different contexts. Said differently, it automatically identifies interactions between variables. The other two techniques, especially logistic regression, attempt to identify a single most-suitable environment as defined by a particular combination of variables. Thus, for many elements, CART may be a more biologically-realistic way to model distribution. For example, assume that amphibian species Z occurs in shaded, cool stream segments at low elevations, and exposed, warm stream segments at high elevations. But at mid-elevations temperatures are just right in all stream segments, and Z's distribution there is driven by stream gradient - it occurs only in relatively flat segments. Given appropriate data on occurrence and predictor variables, a CART model would likely identify the 3 discrete suitable situations (low and cool, mid and flat, high and warm), whereas a logistic regression model would not unless very careful attention was paid to variable interactions, partial contributions, non-linear relationships, residual distributions, and all those other things we are supposed to monitor but never really do.

.

Similar to logistic regression, there are some recently developed programs that facilitate CART-based EDM within the ArcView GIS (http://www.gis.usu.edu/~chrisg/avext/). WYNDD has put together similar programs for moving data between ArcInfo GIS and S-Plus statistical package for CART modeling, and we will gladly move them along to anyone who would like them.

CART also has its analog of requiring the user select a threshold point for the final model. In the CART world, threshold selection it is referred to as "pruning" the output tree. If run in an unsupervised fashion CART will explicitly overfit occurrence data to all environmental variables, and the default output is a very long and complex decision tree that ends in subsets of purely-present and purely-absent points. It is up to the user to prune this tree back several levels to a more parsimonious model that still adequately, but not perfectly, divides present and absent points into discrete sets. The modeler can and probably should set some *a priori* pruning rules in order to produce more objective and defensible models.

There are certain variations of CART that appear to be superior for EDM. In particular, the "random forests" and "bagging trees" approaches are emerging as preferred applications (Prasad et al. 2006). Both of these approaches are iterative, in that they build multiple output trees by continually resampling the presence and absence data and then present an "average" tree as the final output.

4.4.5 INDUCTIVE MODELING - MAXIMUM ENTROPY

Maximum entropy analysis, or "MaxEnt", is based on statistical mechanics methods for making predictions from incomplete information. It evolved from signal processing routines that maximize the signal to noise ratio in a data sample. In the context of EDM, MaxEnt estimates the most uniform distribution (maximum entropy) of occurrence points across the study area given the constraint that the expected value of each environmental variable under this estimated distribution matches its empirical average (Phillips et al. 2004, 2006). The raw output is a probability value (0 - 1) assigned to each map cell in the study area. These values are then converted to a percentage of the cell with the highest probability value, and termed the "cumulative value" in the output map. Thus, threshold selection here is similar to threshold selection in logistic regression.

Comparative studies using MaxEnt in EDM suggest that it is more accurate than others (Phillips et al. 2004) and rather robust to low sample size of occurrence data (P. Hernandez, University of Toronto,

unpublished data). Elith et al. (2006), in their rather exhaustive comparison of several modeling techniques, concluded that MaxEnt is one of the best EDM algorithms available today. And it is similar to DOMAIN in that it does not require absence or pseudo-absence data - the modeler need only bring present data into the MaxEnt routine.

Phillips et al. (2006) place MaxEnt into the same broad category as generalized linear models (GLM), generalized additive models (GAM), and possibly even some Bayesian modeling approaches. Barry and Elith (2006) also note similarities between MaxEnt, GLM, and GAM, specifically in their abilities to fit the complex, nonlinear response surfaces so often found in biological data. One might extend this group to also include the Multivariate Adaptive Regression Splines (MARS) algorithm, which appears to model species distributions better than some other techniques when nonlinear responses are present (Munoz and Felicisimo 2004).

4.4.6 INDUCTIVE MODELING - BAYESIAN TECHNIQUES

We won't pretend to know much about Bayesian statistics, beyond a growing uneasiness that they appear to be seriously challenging traditional parametric (or, more precisely, "frequentist") statistics in natural resource science and will someday force most of us back into the classroom. The basic logic of Bayesian statistics makes a lot of sense in an EDM context, and some applications are being forwarded (e.g., Wintle et al. 2003, Argaez et al. 2005, Gelfand et al. 2006; see also Guisan and Zimmermann 2000). It is perhaps worth noting that one of the first articles to appear in Volume 1- Issue 1 of the new journal "Bayesian Analysis" pertains to EDM (Gelfand et al. 2006).

One contribution of Bayesian statistics may be not as another modeling approach per se, but rather as a method of combining the output from disparate modeling techniques into a synthesized output, thus creating distribution models that draw on the strengths (and mitigate the weaknesses) of the individual techniques. However, one foreseeable problem with such a multi-model combination is that the interpretability of the final model may be diminished as a result. This is discussed in more detail in Section 4.4.7.

Argaez et al. (2005) present a Bayesian approach that integrates expert opinion on a species' distribution, inductive predictions of the species' distribution, and detectability of the species over the study area (i.e., the spatial bias in present- and absent-points resulting from uneven sampling effort). Their approach produces not only a predictive distribution map, but also a confidence map indicating areas for which there is good underlying data, and areas for which there is little or none (see also Gelfand et al 2006).

The Bayesian approach is different enough from deductive modeling and existing inductive techniques (although see Phillips et al. 2006 for a comparison of Bayesian modeling to MaxEnt) to at least be mentioned here. Consider this section as a placeholder for yet-to-be-developed Bayesian techniques in EDM… watch this space…

4.4.7 USING MULTIPLE MODELING TECHNIQUES

There are good opportunities to use multiple modeling techniques to produce a single predictive distribution map for a given element. For example, we have combined envelope modeling and deductive modeling to produce 5-state distribution maps for each of several vertebrates. For each species, after collecting, filtering, and processing occurrence data, we ran a DOMAIN envelope model using elevation and climatic variables to map a general biophysical region of occupation. We then clipped that region by the landcover types identified by experts as suitable for that element, yielding the final distribution map.

We've done similar modeling for some rare plants, using expert-identified soil types rather than landcover types as the clipping layer. In another case we produced a statewide CART model for a riparian-associated vertebrate without using any predictor variables relative to riparian corridors or vegetation (because good riparian data aren't available for Wyoming). We then clipped the CART output with a buffered hydrology layer wherein small streams were buffered by small distances and larger streams were buffered by larger distances, to approximate the location of riparian environments.

Although these examples can be described as using different modeling techniques "in series" (i.e., one is used after the other), they can also be seen as using different techniques "in parallel" (i.e., 2 separate models are produced independently, and then overlain on one another). Instead of describing the first example as first-DOMAIN-then-an-expert-opinion-clip, we could describe it as an independent DOMAIN model intersected with an independent deductive model.

We like the latter description, because it emphasizes that although there are many different modeling algorithms rooted in different logic and statistical families, none is known to be clearly better than the rest in all contexts and circumstances. Some techniques perform better than others under standardized comparisons, but the differences tend to be much less (or even disappear) when real datasets are used (Moisen and Frescino 2002, Elith et al. 2006; also P. Hernandez, University of Toronto, unpublished data). Some techniques (e.g., DOMAIN, MaxEnt) work better with low sample sizes than others (e.g., BIOCLIM). Some (e.g., logistic regression, CART) contrast presence with absence, which can contribute valuable information but requires compilation of absence data, and some (e.g., most envelope models, MaxEnt) do not. Some (e.g., most inductive approaches) are objective and deterministic in result, and some (e.g., deductive approaches) are somewhat subjective and not as replicable. Some may predict well regardless of study area size, whereas others may only perform well across large areas (see Thuiller et al. 2003).

Finally, it is reasonable to assume that some techniques work better for particular species than others (Thuiller 2003, Segurado and Araujo 2004). The distribution of relatively immobile habitat specialists (like some rare plants) may be best modeled by discriminant techniques like CART that identify discrete suitable environments within large and heterogenous areas. Mobile generalists may be better modeled via some of the climatic envelope techniques.

Every EDM project involves a unique combination of data quality, data quantity, study area size, study area history, element biology, intended model use, and other factors. Thus, it is unlikely that the "best" modeling technique can be readily identified in every case (Robertson et al. 2003, Seguardo and Araujo 2004). Given this complexity, the best approach may be to use several different techniques to model the distribution of any given element, and then overlay the models to produce a final map. This approach could be described as "triangulating" on the true distribution of the element, similar to how a navigator triangulates on a geographic feature to estimate its true location. Just as the navigator gets a better estimate of a feature's location by taking multiple bearings, the EDM modeler gets a better estimate of an element's true distribution via multiple models of one set of occurrence data. If three different techniques are used to model and map the same dataset, the modeler should have high confidence in the areas predicted as suitable by all three, and also in the areas predicted as unsuitable by all three. He/ she would have less confidence in areas predicted as suitable or unsuitable by only one or two of the techniques.

There are some canned EDM packages that are built explicitly on this idea of integrating different models into a single output map using far more complex combination routines. In the interest of giving readers a complete picture, and because there have been many papers published using the approach, we'll briefly outline one approach now in circulation: Genetic Algorithm for Rule-set Production (GARP).

GARP - This artificial intelligence approach iteratively applies and recombines different inductive algorithms to occurrence data in a search for an optimal set of models. The current desktop version employs four algorithms: atomic, logistic regression, bioclimatic envelope, and negated bioclimatic envelope rules (Stockwell and Peters 1999). Details of the routine are at http://biodi.sdsc.edu/Doc/GARP/Manual/manual.html. It can be summarized in 6 basic steps:

1. Split occurrence data into modeling and evaluation sets.
2. Model the modeling data set with each of four different algorithms.
3. Translate the initial four models into a common format of "if-then" statements. Thus, each model becomes a series of component if-then's linked by "and's", "or's", or other conjunctions, rather than a series of variable coefficients linked by mathematical operators). The translated models are called rules, and the full set is termed the rule population.
4. Draw a random subset of rules from the population and determine how well each predicts the evaluation data. Throw out those with low predictive success, and retain those with high predictive success and put them back into the population.
5. Draw a new random subset of rules from the population and variously change, join, and recombine their component if-then statements. Put the recombinants back into the population.
6. Go back to step 4; repeat until there is no substantial change in the rule population from one iteration to the next; i.e., the process has converged on a population of consistently accurate rules.

The routine is obviously inspired by biological evolution. Individuals (rules) in a population (rule population) that have the highest fitness (highest predictive accuracy) survive to recombine their genes (component if-then's) to produce a new generation of individuals (new rule population). Differential selection of individuals continues until all individuals in the population have gene combinations that confer high fitness.

Hard-core Darwinists will point out that biological populations rarely achieve such equilibrium because the environment is always changing. GARP approximates this by continually re-sampling occurrence data into different modeling and validation sets, thus always changing the "environment" to which the rules must adapt. Here, though, the biological analogy breaks down - there are a finite number of re-sampling iterations, and the process eventually converges on a single set of well-adapted rules. The output of a single run-to-convergence is an overlay of the predictive distribution maps generated by these final rules.

Random selection in steps 4 and 5, randomness in the recombination rules in step 5, and random re-sampling of occurrence data make GARP output stochastic. Different runs of the same initial data will produce different final rule sets and maps (Anderson et al. 2003). For this reason GARP is often run to convergence many times for the same occurrence dataset. The user can then select a number of the best rules - say the 10 best - from all runs and overlay them to produce a final map with cells having values between 0 (not selected as suitable by any model) and 10 (predicted as suitable by all 10 models).

GARP integrates many EDM steps, and could be discussed as a model calibration technique, a stand-alone modeling algorithm, or a multiple modeling technique (as done here). From the perspective of the end user, however, GARP does not integrate these steps in any straightforward or transparent way. It is a classic "black box" routine. It is not readily apparent how the final output is derived, it is very difficult to communicate the process to an end-user, and the relationship between species occurrence and each environmental predictor is not immediately apparent (Elith 2002, Phillips et al. 2004). Furthermore, formal comparisons of modeling

algorithms appear to regularly identify GARP as one of the least accurate of many algorithms commonly used today (Phillips et al. 2006, Elith et al. 2006; also P. Hernandez, University of Toronto, unpublished data).

The basic idea of combining different models is valuable, but it may be preferable to take a more straightforward approach whereby a set of independent, deterministic, easily-explained models are integrated by a simple map overlay showing areas of concordance and discordance. It may be less sophisticated mathematically, but the result is certainly more easily communicated and understood, and thus has a greater chance of being supported and applied by end-users. It is more similar to the thinking behind the BIOMOD system (Thuiller 2003) than to GARP.

Again, there may also be some value in considering Bayesian combinations of multiple models, whereby one model sets prior probabilities that are subsequently modified by other models. But this too may result in problems of model interpretation and explanation, which is an important consideration in applied ecology. The goal of the applied modeler is not necessarily to produce the best model, in terms of statistical artistry, but rather to produce the best model *that will get used.* Arcane mathematics may produce great models, but they may also predispose those models to being ignored by skeptical users.

The four promising inductive methods outlined above (DOMAIN, CART, multiple logistic regression, and MaxEnt) come from different statistical families, which is important when combining models as suggested in this section. To return to the already overused triangulation analogy, navigators ensure that their independent compass bearings come from widely-separated locations to minimize the size of the plotted error polygon. If the reference positions are too close together they produce a big, funky, stretched-out error polygon. The same thing is true of EDM using multiple models: the separate estimates of distribution should come from widely-dispersed statistical perspectives, lest ye invite more funkiness than necessary.

4.5 MAPPING AND EVALUATION

As mentioned earlier, mapping and evaluation do not necessarily occur in that order. For that matter, with some of the pre-packaged algorithms that integrate right into GIS, even the modeling step is hidden - the first thing the modeler sees is a map. He/ she has to work backward to understand the underlying model, and forward to evaluate it. In all cases, but especially these messy ones, it is helpful to continue thinking about the whole process in terms of the three discrete compartments of modeling, mapping, and evaluation.

4.5.1 MAPPING

Because modeling techniques are so varied, there is no standard procedure for expressing a model in map form. The mapping of deductive models is usually straightforward. Simple models may require GIS identification of only one or two classes on only one environmental layer - e.g., select land cover types 1 and 4 to show the distribution of species Y. More complex deductive models, involving more classes and more layers, can be expressed by intersecting layers - e.g., using these layers on elevation and landcover, select all sites between 3000 - 8000 ft elevation that are within landcover types 1 and 4.

Again, many envelope modeling packages integrate right into a GIS and blend the modeling and mapping steps almost seamlessly. Some of these packages are getting downright spiffy. For example, DOMAIN is available as a menu-driven MS Windows application that produces an ASCII file which can be relatively easily moved into standard GIS platforms, making it (almost) easy enough for even Directors of

state Natural Heritage Programs to use. MaxEnt is similarly accessible. Multiple logistic regression and CART models are becoming easier to map as they are increasingly supported by program sets that tie together GIS and statistical software.

However, as it stands now, there is still a relatively high level of GIS expertise required to map most EDM output. The GIS specialist has to manipulate and combine the relevant environmental layers according to the form and coefficients of the derived model, and apply the model threshold to clearly divide the study area into suitable and unsuitable regions.

Clearly, mapped models need to be delivered in digital form. Most users will end up using them within their own GIS, and hard copy maps have only limited utility. Having said that, it is still appropriate to produce and deliver hard copy maps because that is often the best way to communicate some map limitations and subtleties. Beyond all the standard cartographic conventions of scale bars and north arrows and so forth, there are at least three important ingredients to an EDM hard-copy map:

> *Background:* An adequate number of background features to orient the user - e.g., state boundaries, county boundaries, major roads, major streams. Enough for orientation, but not so much that it clutters up the map.
>
> *Data points:* Points of known occurrence (and points of suspected absence, if used) should be clearly shown, with symbology that distinguishes points in the modeling set from points in validation set. Mapping data points is very important, in that it conveys known distribution that can be contrasted, at a glance, with predicted distribution.
>
> *Model output:* The whole point of the map in the first place. If you stay with semi-transparent color schemes for the model output, the background and point data can show through.

We are still debating whether or not to add a fourth ingredient to this list: prediction accuracy. As discussed earlier it is vitally important to convey that maps produced by EDM are models and not direct representations of element distributions. One way to hammer that home is to stamp every hard copy of a predictive map with its estimated accuracy. Refer to Elith et al. (2002) for a broader discussion of presenting uncertainties in map form.

Depending on the particular element and map, there may be other "accessories" to place on hard copy maps. For example, known range boundaries can help in cases where a model clearly overpredicts by identifying suitable environments in areas where the element is known not to occur (see *The biogeographic problem* in Section 5.0). This may be particularly important when modeling the distribution of an invasive species, in which case it is very important to show where the element is known versus where it might become established.

4.5.2 EVALUATION

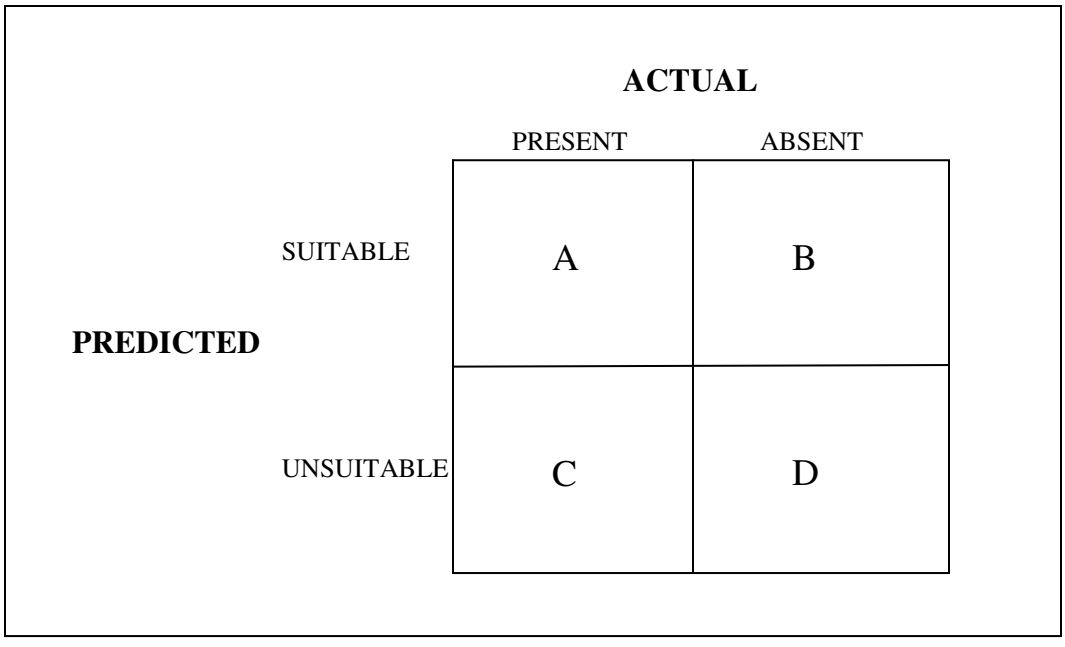
There are as many different ways to evaluate a model as there generating the model in the first place. Like the models themselves, each evaluation method has its own strengths and weaknesses. The one to use for a given model (or set of models) must therefore be selected with the goals of the project in mind.

Biologist review: It's hard to overestimate the value of having a knowledgeable biologist supervise the modeling process and resulting map. There are enough subtleties involved in EDM that it's dangerous to proceed all the way to map delivery without such critique. If nothing else, such critique will help fine-tune the final map so that it conveys better information to the end user.

For example, fishers (*Martes pennanti*) are not known to occupy the Southern Rocky Mountains (which extend into southern Wyoming). However, a by-the-book model of fisher distribution in Wyoming would likely show suitable habitat on the Southern Rocky Mountains, because the subalpine forests there are similar to those on the Central Rocky Mountains where fishers occur. Without a biologist review that notes this overprediction the model and resulting map may be viewed with skepticism by some users - "everyone knows that fishers don't occur there, so why did these yahoos map it there?" Add to this the fact that there are several unconfirmed and unreliable reports of fishers from southern Wyoming that, if used blindly as modeling input, would expand predicted distribution in the area. Again, biologist review of input data would prevent this mistake and raise model validity and credibility among end users.

In short, although EDM can be done without biologist input (especially as canned algorithms and menu-driven GIS routines become more prevalent), it can't be done very well. This argues in favor of a somewhat de-centralized modeling strategy whereby teams of local biologists and GIS specialists (state Natural Heritage programs, perhaps?) produce distribution models for elements with which they are familiar, as opposed to a central-laboratory strategy that relies on canned modeling algorithms, validated solely by statistical procedures, under the direction of personnel with perhaps less knowledge of the study area, input data, and target elements.

Quantitative evaluation: The classic way to evaluate the accuracy of an element distribution model is to challenge the model with independent (i.e., not used to build the model) occurrence data (Rushton et al. 2004). The basic question is how well does the model place independent points of known occurrence within suitable environments (and independent points of suspected absence within unsuitable environments, if appropriate)? Several measures have been used to summarize model success; most are based on different ways of smashing together the values in the quadrants of the basic "confusion matrix" (Fielding and Bell 1997, Guisan and Zimmermann 2000, Pearse and Ferrier 2000b, Manel et al. 2001), which has to be the most honestly-named statistical construct ever. The below diagram and discussion of a confusion matrix is shamelessly pirated from Fielding and Bell (1997):



The quadrants of the matrix (labeled here as A, B, C, and D) are populated by cross-tabulating the actual (or observed) and predicted category of each point in the evaluation set. Clearly, the best models place most points in A and D and the fewest in B and C. There are many ways of combining the numbers in these quadrants for a summary measure of model accuracy. The simplest is probably overall classification rate (or overall classification success), which is:

Overall classification rate = (A + D) / (A + B + C + D)

Substitute (B + C) for (A + D) and you get the mirror image measure, overall misclassification rate (or overall error rate). These overall measures can be broken into present success rate and absent success rate:

Present success rate (or "sensitivity") = A / (A + C)

Absent success rate (or "specificity") = D / (B + D)

These can be easily turned into their mirror image measures of false negative rate (replace A with C in the numerator of the first formula) and false positive rate (replace D with B in the numerator of the bottom formula.

The three summary measures presented above as formulae capture quite a bit of information on model performance, and when presented together they give most users a good sense of model quality. Beware models that are evaluated only by overall classification success, with no indication of present or absent success rates! It is common for poor models to have high overall classification success rates if there are different numbers of present points and absent points in the dataset being evaluated (the ratio of present points to absent points in a dataset is generally referred to as "prevalence" in the EDM literature) (Olden et al. 2002). Imagine a dataset that contains 10 present points and 90 absent points. A model that simply predicts absence at every point in the study area would have an overall classification success of 90%! But it would have a present success rate of 0%.

There are many other evaluation measures which are more robust to the vagaries of real-life data (e.g., outliers, skewed prevalence). Fielding and Bell (1997) discuss the merits and meanings of 13 different summary measures derived from the confusion matrix, and that is by no means a complete list. Without going into detail on all possible measures (some have formulas about a foot long when written out), it is important to recognize that each tends to emphasize a particular aspect of model performance. Thus, each measure serves a different purpose, and with careful consideration of the intended use of a model (guiding field survey? conservation planning? are false positives more detrimental than false negatives, and thus should be weighted more in some sort of cost analysis?) a modeler should be able to select appropriate additional measures to accompany the basic 3 (Fielding and Bell 1997, Guisan and Zimmermann 2000).

Clearly, in order to populate a confusion matrix a modeler must select a threshold value that converts model output into a binary format - suitable vs. unsuitable - and thus threshold selection has a large impact on success rates and other summary measures. There are, however, some threshold-independent measures of model performance, the most common being the receiver operating characteristic (ROC) developed by signal processing and medical researchers. Details of ROC calculation and application are in Fielding and Bell (1997). Briefly, the procedure plots (sensitivity vs. (1 - specificity)) for all possible threshold values. The resulting curve can be used for two purposes: (1) a threshold-independent measure of model performance, which is the area under the curve, and (2) selecting an optimal threshold for mapping suitable and unsuitable environments, which is done by assigning relative weights to false-positive and false-negative cases, combining those weights with the prevalence of positive cases to derive a slope value, then determining the threshold value at which a line with that slope is tangential to the

curve. Blessedly, there are ROC calculators available on the web (e.g., http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html) that are immensely helpful. Most are designed for medical researchers, but with only a little investigation EDM modelers can get their data in the appropriate formats and receive ROC evaluation measures fairly easily.

Depending on which modeling technique is used, quantitative evaluation may occur solely by mathematics. That is, for each point in the evaluation set, the environmental values for the point are entered into the derived statistical model, the model is solved for those values, the result is compared to the suitable/ unsuitable threshold value, and the point is entered into the appropriate quadrant of the confusion matrix. Once all evaluation points have been categorized in the matrix, the overall model success is calculated, and then the final map is generated and delivered.

Alternatively, evaluation may occur geographically. That is, the model is mapped within a GIS, with the threshold point clearly assigning each map cell to the suitable or unsuitable category. The evaluation points are then digitally plotted onto that map and assigned to the appropriate map category. The confusion matrix is built, model success calculated, and the map is delivered. The point here is that mapping of the model came before evaluation, because the map was an integral part of the evaluation process.

Multiple modeling approaches, whereby multiple independent models are overlain spatially to form a final distribution estimate, basically require geographic evaluation because it is not immediately apparent how (or even if) multiple statistical and/ or expert opinion models can be arithmetically combined for purely mathematical evaluation.

There were two EDM evaluation issues raised at the end of Section 4.2.5 that should be mentioned again here. First, some researchers (e.g., Fielding and Bell 1997) suggest formal evaluation of a pre-model, using the independent occurrence data and confusion matrix approach outlined above, and then building a final model using all occurrence data pooled back together. But our previous comment on this still stands: because the measures of model performance are based on the pre-model only, how well do they really represent the performance of the final model?

Second is the issue of evaluation by statistical re-sampling. Most modelers end up bemoaning, "I have a limited amount of hard-earned occurrence data, and my goal is to build the best model possible. So it just doesn't seem right to deliberately reduce my sample size by 25% right off the bat". Conceptually it should be possible to deal with this by re-sampling - that is, for a given algorithm, generate and evaluate many different models based on many different subsets of the occurrence data. In general terms the final model would be an average, either mathematical (e.g., average values for coefficients) or geographical (e.g., number of times each map cell was scored as suitable), of all sub-models, and the accuracy of the final model would be an average of the predictive successes of all sub-models. This approach would take advantage of the full predictive power of all of the occurrence data. But it would also require a lot of computation and computer power (especially for certain algorithms, like CART), and would also require a rather beastly set of custom-built software programs. Such programs are not currently available for EDM right now, but the foaming cauldron of current EDM research and application may spew them forth at any minute.

Field evaluation: Some argue that the best way to evaluate a predictive map is to get out in the field and look for the target element in areas predicted as suitable and areas predicted as unsuitable. This has a lot of appeal, especially since the sampling can be planned in ways that best test the model: equal sampling effort in areas predicted suitable and areas predicted unsuitable, even distribution of sampling effort across the whole study area, field methods standardized by season and time of day, etc. (Vaughan and Ormerod 2003). But it is expensive in time, manpower, and money, so it often goes unsupported by

model requesters. If quantitative evaluation is done right it offers the uneasy consolation of virtual field evaluation. After all, the occurrence data in the evaluation set is field data, and the model was kept blind to that data until the evaluation process… that's sort of like field evaluation … right? …

5.0 SOME OTHER PROBLEMS/ CHALLENGES/ OPPORTUNITIES ENCOUNTERED

EDM for aquatic elements: This is about one generation behind EDM for terrestrial elements, as evidenced by, among other things, the fact that aquatic elements were not included in the first round of USGS Gap Analysis Projects. There are unique challenges to modeling the distribution of aquatic elements. In a general sense, terrestrial taxa sample their environment mostly by moving through it, so we can be reasonably confident that the distribution of individuals reflects the distribution of environments that maximize survival and reproduction. But although aquatic taxa also move, they can also efficiently sample their environment by staying in one place and letting their environment move past them. Thus the distribution of individuals in a stream network may not reflect just the environmental features at those points of occurrence, but also environmental features and qualities up-basin. On top of this, it is hard to find consistent, high-resolution maps of water quality and streambed substrate and other important aquatic features that drive the distribution of aquatic elements.

EDM for riparian elements: The main problem here is that it is hard to find consistent, high-resolution maps of riparian environments that cover large study areas in consistent fashion. Riparian environments tend to occur as thin stringers or small patches that are often below the resolution of mapping projects, so many land cover maps do not show the true extent of riparian environments. The result is that occurrence points for riparian elements often map onto upland types rather than the riparian types in which the observations were actually made, and thus the riparian signal in the element's distribution is degraded. On top of this, most occurrence points have some level of error in mapping precision and thus have a greater tendency to map outside of thin riparian corridors. Again, this masks the actual riparian affinity of the taxon in the modeling phase. It is also a problem in the validation phase. Validation points that come from observations within riparian corridors, but because of low mapping precision map outside of such corridors, will be scored as "misses" by the model when they were actually "hits".

Wyoming is a good example of a state with rather poor riparian data. We have tackled modeling of riparian elements here in a couple of ways that might spark ideas for others in similar situations. First, we've used a multiple modeling approach whereby we intersect a statistical model (e.g., DOMAIN or CART) of distribution, without any riparian or stream network information, with a buffered hydrology layer. This layer, built by the Wyoming Gap Analysis Project (Merrill et al. 1996), is a rough approximation of the extent of riparian vegetation. Small headwater streams are buffered by only a few meters, because riparian vegetation generally occurs there as a narrow strip only a few meters wide. Larger streams, lower in basins, are buffered by successively wider distances, until the largest rivers in the state are buffered by ca. 100m. So the final map shows the buffered stream segments that occur within a generally suitable physical environment for that taxon. It's not perfect, but it's a good estimate in the absence of better data on riparian landcover. Essentially, we are admitting that we are more confident in the ability of the stream layer to show riparian habitats than we are in the ability of the best available landcover layer to do the same.

The other technique we've tried involves using distance-to-stream as a predictor variable in statistical models. Again, this grows out of our assessment that in our area the best streams layer is better at indicating riparian habitat than the best landcover layer. By using distance-to-stream, we don't have to worry about the completeness and accuracy of a riparian habitat map (because we aren't using one) and we don't really have to worry much about the mapping precision of the points of occurrence (within

reason…). As long as the points consistently fall close enough to streams to define a detectable statistical association, this appears to produce good models and resulting maps.

The biogeographic problem (for lack of a better term): In several situations, a good predictive distribution map will predict suitable habitat for a target element in areas where we know, or are pretty darn sure, that that element doesn't occur. For example, predictive distribution maps for pinyon pine (*Pinus edulis*) show suitable habitat in northern Wyoming, 400 mi north of the nearest pinyon stem. Predictive distribution maps for Abert's squirrels (*Sciurus aberti*) similarly show suitable habitat on the Black Hills, 300 mi north of the nearest Abert's squirrel. This is not due to errors or incompleteness in occurrence or environmental data. Rather it is due to historical biogeography, post-Pleistocene dispersal, and regional movement barriers. Put simply, pinyon pine just hasn't made it to northern Wyoming yet. Give it another 1000 years or so and it may get there to occupy those suitable environments. Similarly, Abert's squirrels were isolated to the southwestern U.S. and northern Mexico during the Pleistocene, and have by now dispersed far enough north to occupy extreme southern Wyoming. But there is a gap in ponderosa pine (*Pinus ponderosa*) forest (Abert's squirrels are ponderosa obligates) between the Southern Rocky Mountains and the Black Hills that they have not yet surmounted. Again, given some time, they might get there. Leathwick (1998) and Guisan and Thuiller (2005) discuss distribution modeling of species that are not in equilibrium with their environment.

So it's not a case of mistakenly mapping unsuitable environments as suitable; it's a case of mapping suitable environments that are still waiting to be occupied (a problem of "history of place", as described by Guisan and Zimmermann 2000). How can such biogeographic limits be shown on a predictive map? The most straightforward way is to do just that - show the suspected range boundaries on the final predictive map. In other words, don't mess with the modeling at all; let the model extrapolate where it wants to. But annotate the final map to show suspected biogeographic boundaries, clearly describing in the report text and map caption the reasons why the range boundaries are shown. And if points of known occurrence are shown and delivered with the final predictive map, the user will be further reminded of the difference between known and predicted distribution.

Another technique currently being explored is using distance-to-nearest-present point as a predictor variable in statistical modeling. In other words, deliberately constrain the model extrapolation to just the region around known present points. We are currently researching this technique, and frankly are a little leery of the approach. There are more sophisticated ways, using certain spatial statistics, to constrain the extent of model output relative to the extent of input data (for example, see Argaez et al. 2005, Gelfand et al. 2006).

No matter how it is approached, solving this problem of biogeography requires a bit of ecological artistry. Are you certain enough about the range boundaries to denote them on the map or statistically constrain the model extent, or is there enough reason to suspect presence outside of known range to be cautious about showing such limits? After all, one reason to do EDM is to identify areas outside of currently known range that might support the target element. The first technique of showing suspected range boundaries and points of known occurrence along with predicted distribution conveys a lot of information in that respect. The latter technique of tweaking the modeling process is somewhat dangerous because it will explicitly limit the extent of predictions, and thus probably should be used only for species whose actual ranges are known with confidence.

There are a couple other important issues related to the biogeographic problem. In some cases, identifying suitable habitat outside of the native range of a taxon can be valuable to managers. Fish managers have been deliberately moving game species between watersheds and ecoregions for decades, and managers and sportsman have been accidentally moving bait fish and other species around for at least as long. For example, the upper North Platte River basin in Colorado and Wyoming was thought to be

completely trout-less prior to settlement. Now, after about 100 years of deliberate translocations, it is a major trout fishery. Managers are interested in knowing where particular species and subspecies of trout are most likely to thrive in the North Platte, a question that can be addressed via EDM.

But as valuable as such information can be to managers, it may also be confusing to conservationists. The native range of Yellowstone cutthroat trout (*Oncorhynchus clarki bouvieri*) is the Yellowstone River basin (covering ca. 15% of Wyoming). Yet there is suitable habitat for the taxon in many other basins, and indeed it has been transplanted into most basins of the state. A predictive map showing known occurrences and suitable habitat for Yellowstone cutthroat across the whole state might be misinterpreted in a conservation context: if they are known to be everywhere, and suitable habitat is predicted to occur everywhere, then they can't be a conservation priority, right? There are two major points in the modeling/ mapping process to address here:

> *What occurrence data should go into the model?* Just those points from native populations in the native basin, or points from introduced populations in other basins, too? It is probably most reasonable to base the model on just native points of occurrence because they most directly relate to the environments to which the taxon is best adapted and within which populations can be expected to persist, which is probably the best information to convey to both managers and conservationists. Introduced populations are going to occur where managers put them, not where they evolved, and may persist at those points only because of continued population supplements or other intensive management action.
>
> *How should the resulting map be annotated?* Clearly there needs to be some annotation on the final map. Probably at least (a) the basin boundary of the native range, (b) points of known occurrence used in the model from the native basin, (c) points of known occurrence from outside the native basin, not used in the model, and finally (d) predicted distribution across the whole study area. And, as always, a clear discussion of all the details in the report text and map caption.

This obviously doesn't hold just for fish, but for other taxa that have been moved around deliberately and accidentally. And there are some clear extensions of this discussion to EDM for exotic species.

Taxonomic resolution: This is closely related to the biogeographic problem, and requires yet another visit to Section 4.1. Is the goal of the EDM project to show predicted distribution of a particular population, subspecies, species, or genus (it seems that invertebrate specialists, especially, may be interested in modeling distributions of coarser-than-species taxa)? The answer to this will clearly determine what occurrence data is used as input, and how the final map is annotated.

More importantly, how do we deal with spatial variability in the habitat use and preferences of a given element? Unless they have rather small ranges, most elements probably do not use habitat consistently across their entire range (Dennis et al. 2003). Northern flying squirrels (*Glaucomys sabrinus*), for example, occupy most of northern North America, including the Pacific Northwest, Rocky Mountains, boreal Canada, the upper Midwest, New England, and Appalachian corridor. It is reasonable to assume that the habitat use and distribution of flying squirrels in British Columbia are driven by very different environmental relationships than in North Carolina because the climate, vegetation, soils, flora, fauna, and history of the two areas are very different. If the goal was to model and map distribution across all of North America, then it may be reasonable to include all known points of occurrence, from the Pacific to the Atlantic, in the input data set. But if the goal was to model and map distribution only within British Columbia it would be silly to include points from North Carolina in the modeling data set - the different habits of North Carolina squirrels would seriously mask the relevant patterns of British Columbian squirrels, and result in a poor predictive map.

This is a ridiculously extreme example. But what about including points from Alberta in the input data set for the British Columbia map? How about Montana? Or Wyoming? Including points from nearby areas is a good idea in general, because it increases statistical power by increasing sample size and gives a better sense of suitable environments by increasing the known distribution of the taxon across the ranges of the predictor variables (Van Horne 2002). But the big question is, how far from the study area is too far? Which nearby occurrence points are from populations that use habitat similarly, and thus will help the model, and which ones are from populations that use habitat differently, and thus will confuse the model (a problem of "history of lineage", as described by Guisan and Zimmermann 2000)?

The answer will be element- and area-specific. Some well-studied species have been divided into ecologically relevant subspecies based on differences in habitat use and behavior, and if a target element is one of these the modeler may do well to restrict modeling to just the pertinent subspecies. But intra-species taxonomy is by no means consistent. Perhaps a better rule would be to base distribution models on occurrence data from specific ecoregions, under the assumption that coarse-scale homogeneity in ecological components, processes, and history leads to consistent habitat selection and behavior for a given taxon.

Of course, there are practical constraints to extending a model outside of a particular study area. One common problem is that the spatial layers of predictor variables do not extend in a consistent fashion into adjacent areas. For example, landcover data tends to be state-specific, and it is often difficult or impossible to cross-walk different data sets into a single consistent layer. In these cases occurrence points from two or more states cannot be consistently attributed to landcover, which precludes construction of a single, complete model from regional occurrence data.

There are some ways around this using the multiple modeling approach. For example, to model the distribution of Ute ladies tresses (*Spiranthes diluvialis*) in Wyoming we first built a DOMAIN model of suitable biophysical environments using points of known occurrence from Wyoming and 5 adjacent states (the spatial layers of climate and elevation extended seamlessly across all 6 states). We then we clipped the DOMAIN output with a buffered hydrology layer (this is a riparian obligate plant) and a detailed soils layer (soil affinities of this plant are relatively well-known). The latter 2 layers cover only Wyoming. In this manner we used the out-of-state points to help define the suitable biophysical envelope, then used more detailed in-state layers to refine predictions of occurrence in WY. We have done similar "region-then-state" modeling for other rare elements in the state. Also, Pearson et al. (2002, 2004) have shown success with this approach in modeling distributions of European trees.

Using linear and polygonal analyses: EDM as described in this document assumes that occurrence points are used as input (with some recognition that those points are not mapped with 100% accuracy), and each point is attributed with environmental values at the point location. But there are good reasons to explore the use of polygonal and linear occurrences as input data. After all, most field observations are best mapped as non-point features, given the ubiquity of at least small mapping errors. Furthermore, biologically-relevant syntheses of individual observations (like Element Occurrences, formally) may be the most appropriate input for EDM, and are usually best mapped as polygons. Similarly, a more spatially-explicit analysis of misclassified points during model evaluation may yield more precise estimates of model success.

There are several questions pertinent to using non-point analyses in EDM, phrased below as general research topics:

> *What is the best way to attribute polygonal and linear occurrences with environmental values (means? maximum/ minimum values? variabilty? most common class?)?* Said differently,

what are the best neighborhood expressions of traditional point attributes like landcover, soil type, elevation, etc.?

Is it better to attribute an occurrence with a spatial summary of environmental variables measured across the occurrence's uncertainty polygon, rather attributing it with the environmental values at its centroid? In other words, will spatially-explicit attribution of uncertainty polygons allow coarsely-mapped locality records to contribute more to model performance? Or is it better not to mess with this at all and just throw coarsely-mapped points out of the whole process?

To what extent do "distance-to" environmental variables (e.g., distance-to-stream) differ from neighborhood variables (e.g., stream density within 500m) in contributing to model performance? Neighborhood variables may be more biologically-relevant for some taxa (e.g., relatively mobile vertebrates), but require the selection of an appropriate neighborhood size. Distance-to variables have the advantage of being scale-independent, but may not predict distribution as well.

When cross-tabulating validation points into the confusion matrix, is it appropriate to weight the "misses" (i.e., points that fall in quadrants B and C) by how far (in the literal spatial sense) they are from their correct categorization? Imagine that there are two known-present points in a validation data set that are each placed in unsuitable environments by a model (so they get entered into quadrant C). But the first point maps very close to a map cell scored as suitable, whereas the second point maps way out in the middle of a contiguous block of cells all scored as unsuitable. It seems that the first point is not as much of a miss as the second, and that weighting the misses by distance-to-correct-class could give a better measure of model performance. And what about uncertainty polygons in this context? In some cases a validation point may map onto an incorrect class, but that point's uncertainty polygon overlaps the correct class. Which quadrant of the confusion matrix does that case go into? See Barry and Elith (2006) for more discussion on accounting for spatial patterns in model errors.

6.0 EDM SOFTWARE TOOL

Given all the complexities, unknowns, and options involved in EDM it is difficult for any applied ecology laboratory (a reasonable label for most state Natural Heritage programs?) to apply it without a significant investment of time, money, and expertise. One of the main reasons for producing this document was to help minimize this up-front investment. To the same end WYNDD and NatureServe are creating a software tool that will guide modelers through the whole EDM process and automate many of the complex data processing, modeling, and mapping steps.

This tool will be distributed as Version 1.0 in summer 2006. This version will guide modelers through EDM using 3 algorithms: DOMAIN, multiple logistic regression, and Maximum Entropy. Importantly, the tool will be continually updated and improved to stay current with new research results and modeling approaches. In this manner it will continue to allow users to produce quality models and maps while minimizing the investment required to stay at the leading edge of technical literature, technological innovations, and conceptual advances.

Ultimately the tool will implement a multiple modeling approach that allows modelers to quickly apply several modeling techniques to particular occurrence data sets, then integrate the output from all approaches in summary maps with accurate and relevant evaluation measures. EDM is such a powerful approach to extracting information from raw occurrence data that we envision it becoming a standard

practice within state Natural Heritage programs. We see the production, distribution, and maintenance of this tool, and this document, as a first step in that direction.

LITERATURE CITED

Anderson, R.P. 2003. Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. Journal of Biogeography 30:591-605.

Anderson, R.P., D. Lew, and A.T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling 162:211–232.

Argaez, J.A., J.A. Christen, M. Nakamura, and J. Soberon. 2005. Prediction of potential areas of species distributions based on presence-only data. Environmental and Ecological Statistics 12:27–44.

Augustin, N.H., M.A. Mugglestone, and S.T. Buckland. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology 33: 339-347.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101-118.

Barbour, M.G., J.H. Burk, and W.D. Pitts. 1987. Terrestrial plant ecology. Benjamin/ Cummings Publishing Coimpany Inc. Menlo Park, California, USA.

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

Beauvais, G.P. and R. Smith. 2003. Model of the breeding habitat of the Mountain Plover (*Charadrius montanus*). Western North American Naturalist 63:88-96.

Beauvais, G.P., D.A. Keinath, P. Hernandez, L. Master, and R. Thurston. 2004. Element distribution modeling: a primer -or- Musings upon depicting the geographic arrangement of biota, with ample reference to the harsh lessons of the dark master experience. Version 1.0. Wyoming Natural Diversity Database - University of Wyoming. Laramie, Wyoming, USA.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and regression trees. Wadsworth International Group. Belmont, California, USA.

Carpenter, G., A.N. Gillson, and J. Witmer. 1993. DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation 2:667-680.

Colinvaux, P. 1986. Ecology. John Wiley and Sons Inc. New York, New York, USA.

Corsi, F., J. de Leeuw, and A.K. Skidmore. 2000. Modelling species distribution with GIS. Pages 389-434 *in* L. Boitani and T.K. Fuller (editors). Research techniques in animal ecology; controversies and consequences. Columbia University Press, New York, New York, USA.

Cumming, G.S. 2000. Using between-model comparisons to fine-tune linear models of species ranges. Journal of Biogeography 27:441-455.

Davis, F.W., D.M. Stoms, A.D. Hollander, K.A. Thomas, P.A. Stine, D. Odion, M.I. Borchert, J.H. Thorne, M.V. Gray, R.E. Walker, K. Warner, and J. Graae. 1998. The California Gap Analysis Project--Final Report. University of California, Santa Barbara, California, USA.

Dennis, R.L.H., T.G. Shreeve, and H. Van Dyck. 2003. Towards a functional resource-based concept for habitat: a butterfly biology viewpoint. Oikos 102:417-426.

D'heygere, T., P.L.M. Goethals, and N. DePauw. 2003. Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic invertebrates. Ecological Modelling 160:291-300.

Elith, J. 2002. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. Pages 39-58 in S. Ferson and M. Burgman (editors). Quantitative methods for conservation Biology. Springer-Verlag, New York, New York, USA.

Elith, J., and M.A. Burgman. 2003. Habitat models for PVA. Pages 203-235 in C.A. Brigham and M.W. Schwartz (editors). Population viability in plants: conservation, management and modeling of rare plants. Springer-Verlag, New York, New York, USA.

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

Elith, J., C.H. Graham, R.P. Anderson, M. Dudık, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A.T. Peterson, S.J. Phillips, K.S. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberon, S. Williams, M.S. Wisz, and N.E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.

Engler, R., A. Guisan, and I. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41:263-274.

Farber, O., and R. Kadmon. 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecological Modelling 160:115-130.

Ferrier, S., G. Watson, J. Pearce, and M. Drielsma. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales I: species-level modelling. Biodiversity and Conservation 11:2275-2307.

Fertig, W., R. Thurston, and W.A. Reiners. 2003a. Modeling the potential distribution of plant species in Wyoming: a tool for gap analysis. Wyoming Natural Diversity Database, University of Wyoming, Laramie, Wyoming.

Fertig, W., R. Thurston, and W.A. Reiners. 2003b. Modeling the potential distribution of BLM Sensitive and USFWS Threatened and Endangered plant species in Wyoming. Wyoming Natural Diversity Database, University of Wyoming, Laramie, Wyoming.

Fielding, A.H., and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

Fielding, A.H. and P.F. Haworth. 1995. Testing the generality of bird-habitat models. Conservation Biology 9:1466-1481.

Gelfand, A.E., J.A. Silander Jr., S. Wu, A. Latimer, P.O. Lewis, A.G. Rebelo, and M. Holder. 2006. Explaining species distribution patterns through hierarchical modeling. Bayesian Analysis 1:41-92.

Godown, M.E., and A.T. Peterson. 2000. Preliminary distributional analysis of US endangered bird species. Biodiversity and Conservation 9:1313-1322.

Greaves, R.K., R.A. Sanderson, and S.P. Rushton. 2006. Predicting species occurrence using information-theoretic approaches and significance testing: an example of dormouse distribution in Cumbria, UK. Biological Conservation 130:238-250.

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 Modeling 135:147-186.

Hosmer, D.W. Jr., and S. Lemeshow. 1989. Applied logistic regression. Wiley. New York, New York, USA.

Johnson, C.J., D.R. Seip, and M.S. Boyce. 2004. A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. Journal of Applied Ecology 41: 238-251.

Kautz, R.S. and J.A. Cox. 2001. Strategic habitats for biodiversity conservation in Florida. Conservation Biology 15:55-77.

Leathwick, J.R. 1998. Are New Zealand's Nothofagus species in equilibrium with their environment? Journal of Vegetation Science 9:719-732.

Loiselle, B.A., C.A. Howell, C.H. Graham, J.M. Goerck, T. Brooks, K.G. Smith, and P.H. Williams. 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology 17: 1591-1600.

MacKenzie D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, L.L. Bailey, and J.E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier Inc. Burlington, Massachusetts, USA.

Manel, S.H., C. Williams, and S.J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38:921-931.

McCarthy, M.A. and J. Elith. 2002. Species mapping for conservation. GAP Analysis Program Bulletin 11:29-31.

Merrill, E.H., T.W. Kohley, M.E. Herdendorf, W.A. Reiners, K.L. Driese, R.W. Marrs, and S.H. Anderson. 1996. The Wyoming Gap Analysis Project final report. Wyoming Cooperative Fish and Wildlife Research Unit. University of Wyoming, Laramie, Wyoming.

Moisen, G.G. and T.S. Frescino. 2002. Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling 157:209-225.

Morrison, M.L., B.G. Marcot, and R.W. Mannan. 1992. Wildlife-habitat relationships: concepts and applications. University of Wisconsin Press. Madison, Wisconsin. USA.

Munoz, J. and A.M. Felicisimo. 2004. Comparison of statistical methods commonly used in predictive modelling. Journal of Vegetation Science 15:285-292.

Olden, J.D., D.A. Jackson, and P.R. Peres-Neto. 2002. Predictive models of fish species distributions: a note on proper validation and chance predictions. Transactions of the American Fisheries Society 131:329–336.

Pearce, J. and S. Ferrier. 2000a. An evaluation of alternative algorithms for fitting species distribution models developed using logistic regression. Ecological Modelling 128:127-147.

Pearce, J. and S. Ferrier. 2000b. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133:225-245.

Pearce, J.L., K. Cherry, M. Drielsma, S. Ferrier, and G. Whish. 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. Journal of Applied Ecology 38:412-424.

Pearson, R.G., T.P. Dawson, P.M. Berry, and P.A. Harrison. 2002. SPECIES: a spatial evaluation of climate impact on the envelope of species. Ecological Modeling 154:289–300.

Pearson, R.G., T.E. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27:285–298.

Peterson, A.T., M.A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberon, R.H. Buddemeler, and D.R.B. Stockwell. 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416: 626-629.

Phillips, S.J., M. Dudik, and R.E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Proceedings of the Twenty-first Century International Conference on Machine Learning.

Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259.

Prasad, A.M., L.R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181-199.

Pulliam, R. 2000. On the relationship between niche and distribution. Ecology Letters 3:349-361.

Reese, G.C., K.R. Wilson, J.A. Hoeting, and C.H. Flather. 2005. Factors affecting species distribution predictions: a simulation modeling experiment. Ecological Applications 15:554–564.

Robertson, M.P., C.I. Peter, M.H. Villet, and B.S. Ripley. 2003. Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. Ecological Modelling 164:153-167.

Rushton, S.P., S.J. Ormerod, and G. Kerby. 2004. New paradigms for modelling species distributions? Journal of Applied Ecology 41:193-200.

Scott, J.M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T.C. Edwards, J. Ulliman, and R. G. Wright. 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildlife Monographs 123:1-41.

Scott, J.M., P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson (editors). 2002. Predicting species occurrences: issues of accuracy and scale. Island Press. Washington DC, USA.

Segurado, P. and M.B. Araujo 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31:1555–1568.

Seoane, J., J. Bustamente, and R. Diaz-Delgado. 2005. Effect of expert opinion on the predictive ability of environmental models of bird distribution. Conservation Biology 19:512-522.

Short, H.L., J.B. Hestbeck, and R.W. Tiner. 1996. Ecosearch: a new paradigm for evaluating the utility of wildlife habitat. Pages 569-594 *in* R.M. DeGraaf and R.I. Miller (editors). Conservation of faunal diversity in forested landscapes. Chapman and Hall, London, UK.

Smith, P.A. 1994. Autocorrelation in logistic regression modelling of species' distributions. Global Ecology and Biogeography Letters 4:47-61.

Southwest ReGAP 2005. Predicted animal habitat distributions and species richness. Provisional final draft report.

Stephens, P.A., S.W. Buskirk, G.D. Hayward, and C. Martinez del Rio. 2005. Information theory and hypothesis testing: a call for pluralism. Journal of Applied Ecology 42:4-12.

Stockwell, D., and D. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science 13:143-158.

Thuiller, W. 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology 9:1353-1362.

Thuiller, W., M.B. Araujo, and S. Lavorel, S. 2003. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. Journal of Vegetation Science 14:669-680.

USDI Fish and Wildlife Service. 1981. Standards for the development of habitat suitability index models. United States Fish and Wildlife Service. Release No. 1-81, 103 ESM.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. Journal of Wildlife Management 47:893-901.

Van Horne, B. 2002. Approaches to habitat modelling: the tensions between pattern and process and between specificity and generality. Pages 63-72 in J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson (editors). Predicting species occurrences: issues of accuracy and scale. Island Press. Covelo, California, USA.

Vaughan, I.P. and S.J. Ormerod. 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conservation Biology 17:1601-1611.

Verbyla, D.L. and J.A. Litvaitis. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. Environmental Management 13:783-787.

Vojta, C.D. 2005. Old dog, new tricks: innovations with presence-absence information. Journal of Wildlife Management 69:845-848.

Williams, B.K. 1996. Assesment of accuracy in the mapping of vertebrate biodiversity. Journal of Environmental Management 47:269-282.

Wintle, B.A., M.A. McCarthy, C.T. Volinsky, and P.R. Kavanagh. 2003. The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology 17:1579-1590.

Zaniewski, A.E., A. Lehmann, and J.M. Overton. 2002. Predicting species spatial distributions using presence - only data: a case study of native New Zealand ferns. Ecological Modelling 157:261-280.

GLOSSARY

Distribution – The environments occupied by an element. Can be expressed at any geographic scale, and can emphasize probabilities of occurrence or presence/ absence. Methods for mapping distributions vary, but are typically spatial extrapolations of models of suitable and unsuitable environments based on known areas of occurrence. Distribution maps, as compared to range maps, depict within-range variation in occupation rather than simply the outer limits of occupied area. See Range, Habitat, Environment.

Element - A term introduced by state Natural Heritage programs to help organize biological information. An element is any meaningful biological unit. It is similar to a "taxon", except an element can exist on the ecological hierarchy (e.g., ecosystem, community, guild) as well the standard taxonomic hierarchy (e.g., Genus, Species, Subspecies). In practice, most zoological elements refer to species or sub-species, most botanical elements refer to species or varieties, and most ecological elements refer to vegetation communities. See Species, Taxon.

Environment - The entire collection of biotic and abiotic features and conditions at a given site. If the collection promotes occupancy by a given element, then that environment is part of the element's range and distribution. If the collection promotes not only occupancy but also reproduction and survival, then that environment is also habitat for that element. See Range, Distribution, Habitat.

Evaluation - In the context of EDM: a process of ascertaining the accuracy or validity of a given distribution model. It can be done via biologist review, quantitative statistics, or targeted field survey.

Habitat – An environment with the combination of resources and conditions that promote occupancy by individuals of a given species (or population) and allows those individuals to survive and reproduce (Morrison et al. 1992). Habitat is therefore a subset of distribution (and distribution is a subset of range), because distribution emphasizes occupancy without any direct reference to survival or reproduction (although survival and reproduction are inferred by consistent occupation; see Range, Distribution, Environment).

Mapping - In the context of EDM: expressing an element distribution model in a spatially-explicit form referenced to real geography, typically performed via geographic information system (GIS).

Modeling - In the context of EDM: generating a verbal or statistical relationship between an element and the environment in a way that defines that element's distribution. Modeling typically uses knowledge of the biology of the element in combination with points of known occurrence to identify environmental variables, and the values of those variables, that discriminate suitable from unsuitable environments (or define categories of probability of occurrence). Distribution models are necessary precursors to element distribution mapping.

Occurrence - In the general context of EDM: a location where a particular element has been observed. Can be expressed as a point, line or polygon, and typically is associated with some degree of mapping error. "Occurrence data" therefore refers to a set of locations where a given

element has been documented. In this general form, "occurrence" is similar to "source feature" (below).

In the context of state Natural Heritage program methodology: an area an area of land and/or water in which a species or natural community is, or was, present. An occurrence should have practical conservation value for the element as evidenced by potential continued (or historical) presence and/or regular recurrence at a given location. For species elements, the occurrence often corresponds with the local population, but when appropriate may be a portion of a population (*e.g.*, long distance dispersers) or a group of nearby populations (*e.g.*, metapopulation). For community elements, the occurrence may represent a stand or patch of a natural community, or a cluster of stands or patches of a natural community. "Occurrence" is often used as a contraction of the more formal term "Element Occurrence". In contrast to the more general definition above, an Element Occurrence is typically not based on a single observation but rather on multiple and repeated observations of the target element at a given site. See Source feature.

Range – The total areal extent occupied by an element. Typically expressed at coarse geographic scales (e.g., continental, regional), and typically emphasize presence/ absence rather than probabilities of occurrence. Methods for mapping ranges vary, can include substantial subjectivity, and are commonly not well-documented. Most range maps are based on simple polygons that encompass the outermost points of known occurrence of an element, and thus do not show much within-range variation in occupation. See Distribution, Habitat, Environment.

Source feature - A mapped observation of a given element. Source features can be points, lines, or polygons, and are typically associated with some degree of mapping error. Clusters of source features are often combined into a single Element Occurrence. See Occurrence.

Species - Keeping in mind that biologists still can't agree on a single best definition: a group of related populations, the members of which compete more with their own kind than with members of other species ("ecological species"; Colinvaux 1986). Or: A group of morphologically and ecologically similar natural populations that may or may not be interbreeding but are reproductively isolated from other such groups ("taxonomic species"; Barbour et al. 1987). See Taxon, Element.

Taxon - A distinguishable unit or level on the taxonomic hierarchy (Kingdom, Phylum → Genus, Species). Typically extends down the standard hierarchy to encompass subspecies, variations, races, and even metapopulations and populations. See Species, Element.