Developing an infrastructure for knowledge based, spatially explicit population modeling of multiple species

Ed Laurent*, Ashton Drew, and Jaime Collazo

Biodiversity and Spatial Information Center & North Carolina Cooperative Fish and Wildlife Unit North Carolina State University, Raleigh, NC 27695-7617;
*Ed_Laurent@ncsu.edu

Anthropogenic changes to vegetation communities, climate, and landscape patterns affect wildlife species' distributions, abundances and population viability. Hence, there is a need to better understand these relationships and predict their consequences given various change scenarios. The development and application of data-driven predictive modeling approaches are limited, however, by the complexity of relationships between wildlife populations and their environments, and by the resources necessary to strategically sample and monitor this complexity over large areas for multiple species. Knowledge-driven modeling, whereby model structure and parameter values are based on a thorough review of the literature, offers a prudent strategy for predicting and mapping population rates. These models can be tested with data when available and be extended to predict changes under various scenarios. This knowledge-based approach may be optimized through an infrastructure that promotes collaboration and innovation via online access to information and tools.

We present an infrastructure framework for predicting species' population rates over their entire geographic ranges and iteratively improving upon those predictions as new knowledge, tools and data become available. Our approach includes three primary environments:

1) Knowledge Environment (collaboration and dissemination)

2) Modeling Environment (prediction)

3) Information Environment (validation and inference).

The Knowledge Environment enables collaboration among unlimited partners for summarizing and disseminating knowledge, tools, and information regarding species' habitat associations, densities, and vital population rates under spatially explicit conditions via Internet services. The Modeling Environment enables partners to design and share components of hierarchically structured, spatially and temporally explicit models of species' distributions and productivity. The Information Environment provides a data structure for validation of predictions in ways that enable feedback to the Knowledge Environment.

As an example of synthesis and synergy among the three infrastructure environments: text descriptions of a species' productivity in a specified spatial context may be summarized online using forms that allow partners to describe knowledge in a standard format. This standardization allows knowledge entries added by various partners to be queried and quantitatively summarized in order to derive parameter estimates of vital population rates and their statistical distributions under spatially explicit conditions. The resulting spatially and temporally explicit parameter estimates and their statistical distributions can be used as informative priors in hierarchical empirical Bayes models that estimate the productivity of a specified area. Such models may be developed using functions pulled from libraries or packages contributed by other partners, or by writing new compatible functions that are later shared. Data can be collected or acquired from partners to evaluate the models, and multi-model inference can be employed for confirmatory analysis of competing model structures. Finally, the process and results of these analyses can be documented and used to inform new model structures, code new models, create multi-author species' wikis, provide guidance to other partners through web forums, and published in the peer-reviewed literature. Combined, this infrastructure framework enables partners to increase their individual output and the efficiency of collaboration. Entropy is reduced across partners by structuring individual actions within a broader, hierarchical context that removes redundancies, facilitates innovation, and enables the sharing of knowledge.

Knowledge environment: **Websites, software, and interactive user interfaces that not only allow data entry, but also encourage it through social networking strategies that distribute the time and effort of summarizing existing knowledge over multiple users.**

Examples employ schema built on Freebase.com that did not require any user programming.

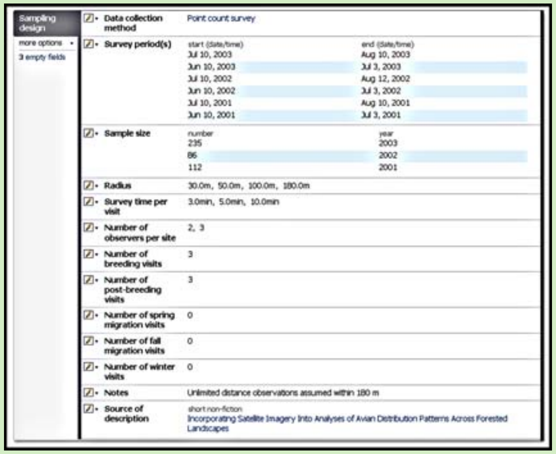

Figure 1. Data collection strategies of published studies are consistently described so that compatible results can be identified and combined for meta-analysis.

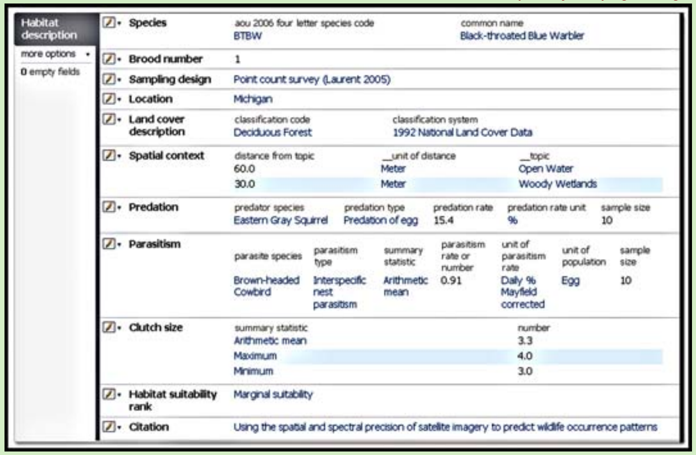

Figure 2. Research results are described using a standard schema so that new knowledge can be can be described within standardized framework, then queried and statistically summarized (fictional data shown).

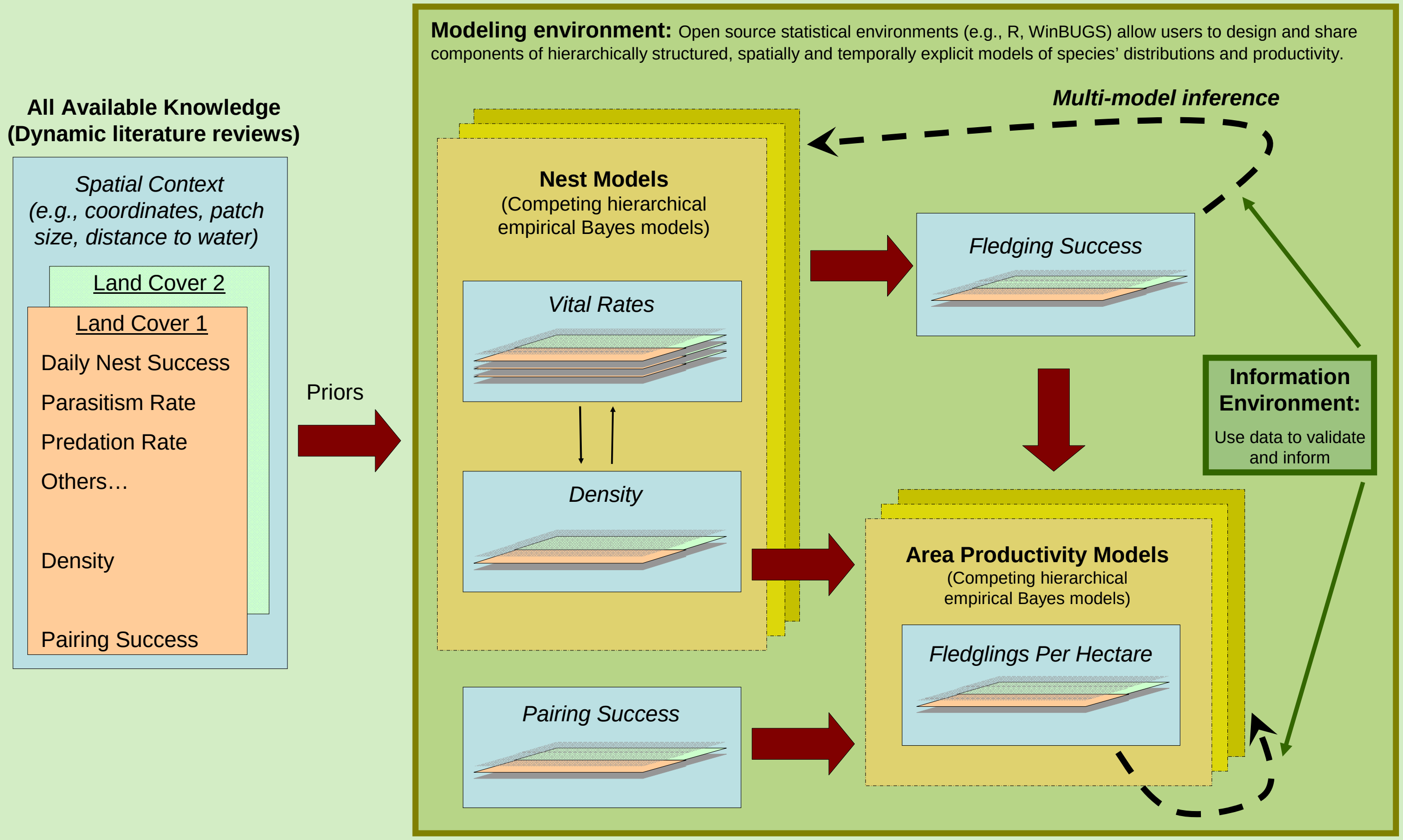

Figure 3. Conceptual model of how spatially explicit knowledge of population rates and density estimates observed within different land cover descriptions and different spatial contexts can be used as informative priors in competing hierarchical empirical Bayesian models that are tested using validation data collected in the field.