

USING INTEGRATED POPULATION MODELLING IN DECISION-SUPPORT TOOLS TO CONNECT SCIENCE AND DECISION MAKERS

A proposal for USGS Mendenhall Research Fellowship Program RO #S31
Bridging the Disconnects Between Science and Decision Making

Jessica L. Burnett, M.Sc.

0 Summary

“...most models in ecology are intended for other ecologists, cooked up by scientists for consumption by scientists...” —Roughgarden (1997)

Herein I propose a project which seeks to strengthen the research-implementation connection, or the bridge between scientists and decision makers [Figure 1(a-b)], by building a high-value decision-support tool (DST) using a powerful yet under-rated, integrative modelling framework: Integrated Modelling. The aim of this project is to identify and ameliorate the barriers to the adoption and use of Integrated Models (IMs) in decision-making processes at various management levels. Adopting these and other integrative approaches to ecological understanding and management becomes increasingly important to both the scientist and the decision-maker as we progress through the era of Big Data [2, 3] and information overload [4]. Refining both the analytical methods and the delivery of IMs will strengthen the bridge between the scientist and decision-maker. This project will not only advance Open Science by encouraging the reuse and integration of disparate, publicly-funded data to advantage scientific understanding, but will link this understanding to a high-value decision support tool for wildlife risk assessment and management.

1 Background and rationale

Decision support tools (DST) are a powerful package within which scientist and modelers can deliver science-based knowledge to decision makers and other end-users. DSTs can be described in two inversely related dimensions: 1) ease of use and 2) utility. Tools that are relatively easy to use and interpret are considered *high-value* DSTs. One way to enhance the value of a DST is to create a tool which incorporates a multitude of information (data), or data gathered from multiple sources. Developing high-value DSTs advances the mission of the USGS by 1) increasing the value of its monitoring programs and data collection schemes and 2) expediting the decision-making and management processes by consolidating information packages regarding the same issue. High-value and real-time DSTs were recently identified as potential solutions or products for the majority of the USGS’s Grand Challenges [5].

Integrated Models (IMs) are a potentially powerful approach for ecological understanding and management and for the decision-making process. Succinctly described, IMs use multiple data or information sources to inform a single model. IMs broadly describe models which incorporate many data sources and types, and have been used to inform decisions ranging from single species (e.g. Integrated Population Models; [6, 7] to whole-system and natural resource management (e.g. Multi-resource Analysis[8]. Integrated Population Modelling has great potential as a high-value DST in applied ecology. Integrated Population Models (IPMs) have,

however, not been adopted by the broader ecological research community, which may be stunting its development as a DST.

Ecological models rarely mature to high-value DSTs when they are fostered only within disciplinary bubbles. Although progress has been made over the last two decades, many powerful and potentially highly-useful tools and knowledge remain unseen by decision makers and consequently do not fully benefit of ecological management, conservation, and science.

Integrated Population Models (IPMs) are an example a potentially high-value tool which is stuck in its disciplinary bubble. Operating within an Open Science framework, I propose to translate IPMs into a high-value DST by incorporating spatial landscape features into existing models. Additionally, I will make an option to conduct spatially-explicit analyses, which allow for spatiotemporal extrapolation by the end-user. These IPMs and DSTs can be used to understand the impact of various management strategies and effects of land use change on wildlife populations at management-relevant scales.

2 Research objectives

This project has three research objectives (Figure 2), each of which can be considered mutually exclusive but will, ideally, interact to build a cohesive decision-support tool (DST; Objective 3).

3.2 2.1 Objective 1: Incorporate landscape features and habitat relationships into the spatially implicit IPM

This objective (Figure 1) incorporates landscape and habitat features and relationships into existing, spatially-implicit IPMs. This objective is a logical first step to this project because identifying and incorporating the appropriate USGS data and models into an IPM will require a significant amount of time and correspondence with multiple collaborators.

3.3 2.2 Objective 2: Incorporate spatial dynamics into the IPM

The model produced in this Objective can be programmed with ease when landscape features data are identified. This Objective incorporates spatial location of sampling sites into the model built in Objective 1 (see 2.1), allowing for more detailed understanding of the influence of spatial variation on population trends and demography.

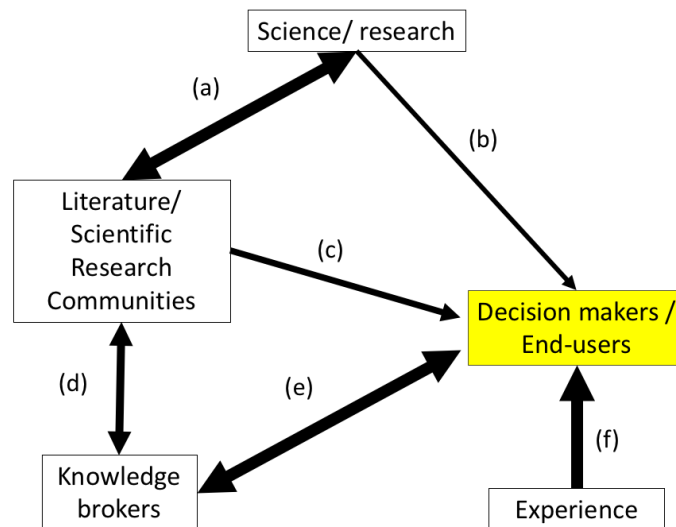


Figure 1. Conceptual diagram of the information and knowledge used to inform decision makers and end-users in applied ecology. Line width, relative strength of connections. Connections (b-c), weak bridges between science/knowledge generation and decision makers. Links (a-d) comprise science-based knowledge transfer, where (e-f) may be further subjected to misinterpretation and noise. Although all connections are important, this project stresses the value of vectors (b-c) in strengthening bridges between science and decision makers.

3.4 2.3 Objective 3: Generate a high-value, generalizable Decision Support Tool

Although the product for this objective is largely focused on programming and software development, a DSTs cannot be of high-value without close collaboration with experts in decision science and implementation, and with end-users and decision makers. I will work closely with Dr. Jenni to identify the barriers common to successful implementation of these types of DSTs. Perhaps as important, I will build DSTs using an agile software development approach, one which seeks iterative feedback from end-users, to improve the accessibility and relevance of the DST while exploiting the benefits of IPMs.

3 Methods

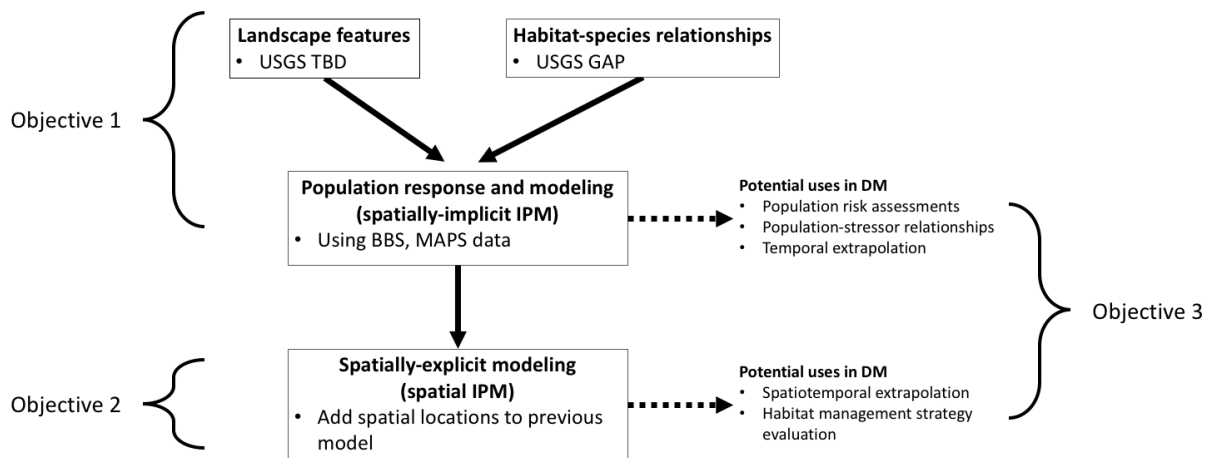


Figure 2. Outline of research objectives, major modelling components, and potential data.

3.5

3.6 3.1 Data

Three broad categories of data are required to achieve our objectives: count data, mark-recapture data, and landscape features. Our IPM (2.1, 2.2) can handle a variety of data and can be adapted to model population dynamics of various species, and because our goal is to create reproducible and user-friendly DSTs, we will conduct a case study for a single species while creating a generic analysis. Significant progress has been made in IPM development using avian populations of North America [9–11], partly due to the successful and long-running avian monitoring program, USGS Breeding Bird Survey (BBS; Sauer et al. 1966). Building off this work, we will build IPMs using birds.

3.1.1 Count data (abundance and occurrence).

We will primarily use species population estimates and occurrence from the BBS. Although numerous monitoring programs exist to obtain better estimates of waterfowl species, these data are not as easily obtained as BBS is, currently. Therefore, we will build IPM for non-waterfowl, breeding birds using the BBS data. BBS data comprise abundance indices and therefore do not allow for estimation of vital rate parameters (see 3.1.2).

3.1.2 Mark-recapture (demography).

Few mark-recapture databases that are available for public use are as comprehensive and widespread as that of the Mapping Avian Productivity and Survival (MAPS). MAPS data is available for many breeding, non-waterfowl birds in North America. The MAPS program was designed to complement the BBS data in that individual-level observations allow for estimation of vital rates.

3.1.3 Landscape features.

The count (BBS) and demographic (MAPS) data provide spatial locations, dates, and times of surveys, providing an opportunity to predict and further extrapolate predictions to unsampled areas of similar landscape and climatic conditions. The USGS is an ideal institution within which to seek landscape features data. I will work with researchers within CSAS&L and beyond to identify the most-appropriate data to use for our IPMs.

3.7

3.8 3.2 Integrated Population Models (IPMs)

Integrated population models (IPMs) are a framework which uses multiple, often disparate, datasets to inform models of wildlife populations [9, 13, 14]. Although a seemingly powerful tool for wildlife conservation and research, IPMs are largely restricted to academic exercises which advance the technical aspects of the methods rather than its contribution to DM processes (Schaub and Abadi 2011; but see Coates et al. 2018, Ross et al. 2018). IPMs are distinct from traditional population models in four ways. First, the framework allows for estimation of latent parameters. Second, combining demographic and abundance data improve precision of parameter estimates. Improving parameter estimates is especially useful when one or more of the datasets in question are of low quality. Third, incorporating multiple data types allows us to account for variation in the observation process (e.g., recapture rate, observer experience) among relevant stages (e.g., age, sex, season). Finally, the IPM framework allows the analyst to combine multiple existing models.

3.1.4 Building the base IPM

The choice of species (and therefore existing IPM) will be restricted by the availability of the appropriate landscape features data and overlap with BBS and MAPS information. This should not, however, be an issue given the plethora of earth-systems data which is publicly available and compatible with our potential study extent and resolution. Candidate IPMs include [7, 16].

Integrated population models comprise, crudely, two sub-models: a state-space (count model) and an

individual-based, demographic (mark-recapture model) model (see Figure 3). State-space models comprise both process and observation components (Figure 3), ideal for accounting for biases and error inherent in human censusing (imperfect detection) of fauna. Stage-structured population models are extremely well-developed, but collecting these data requires a significant amount of resources (time, manpower, and money). IPMs provide a way to incorporate heterogeneous data, allowing us to estimate population size or trends based on both demographic and occurrence or abundance data [10, 13]. We will use a discrete time (annual), stage-structured model (two stages: hatch-year and after hatch-year) to estimate N (true abundance) and ϕ (apparent survival) (Figure 3). Operating within a Bayesian inference framework and assuming a first-order Markov process, we will fit the model using MCMC methods.

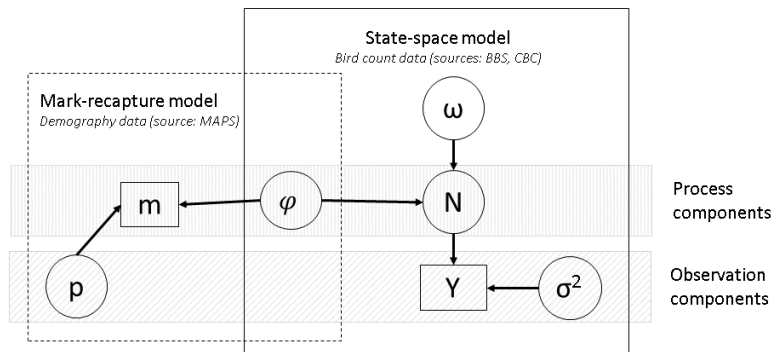


Figure 3. Crude directed acyclic diagram depicting the IPM model structure for using the count and demography datasets. Boxes = observations, circles = estimated parameters. Here, p = recapture probability; ϕ = apparent survival rate; m = individual-based observations; σ^2 = observation error; ω = immigration; N = true population abundance; Y = observed population abundance.

3.1.5 Incorporating landscape features into the spatially-implicit IPM

Landscape features, primarily land cover and land use, will be incorporated into the base IPM, informing the observer component of the state-space model (informing σ^2) and the latent variable, apparent survival (ϕ ; Figure 3). Here, landscape features can be associated, at various spatial scales (e.g., via buffering), with unique sites and observations yet remain unassociated with spatial location.

3.1.6 Making the model spatially-explicit

Adding a spatially-explicit component will require only minor adjustments to the statistical programming code but require significant subject expertise to inform the biological processes associated with spatial variation. A spatially-explicit IPM will allow the end-user to extrapolate predictions to previously-unsampled locations. Unlike the spatially-implicit IPM the spatially-explicit IPM does not require the mark-recapture and count data to be independent [17].

3.1.7 Decision-support Tools

I will seek advice from collaborators with expertise in the decision and translational sciences to identify the science/information needs of our audiences prior to building the IPM (Objectives 1; section 3.2.1). DSTs will comprise R packages, designed for the analyst to have complete control over the Modelling, and a web-based tool (e.g., a Shiny App) designed to allow the end-user to interact with the R package in a web browser with no programming prerequisites.

4 Expected contributions to USGS, CSS and CSAS&L missions

The integrative nature of this proposed work, coupled with my enthusiasm for conducting integrative, relevant, and collaborative science, makes the CSAS&L and the Mendenhall RO #S31 research team an ideal environment in which to succeed. The USGS is uniquely poised to contribute its large-scale, longitudinal geophysical, climatic, and ecological models into existing IPM frameworks. The institutional knowledge within the CSAS&L team will greatly benefit Objective 1, identifying and using the landscape and habitat data products.

This project is low-risk but has the potential to inform decision-makers and practitioners in the short- and long-terms. We will build off of existing IPMs to efficiently incorporate federally-funded land features data into science-based knowledge for end-users, therefore, the success of the project lies mostly in working with USGS and other data product experts to identify and incorporate the best data products into our models. This project contributes to the CSAS&L's Five Year Science Strategy by 'Converting Data into Actionable Knowledge'. This project also aligns well with the CSAS&L Mission, contributing to primarily to the Biodiversity Science theme's Strategy Goals 1 (leadership in biodiversity science), 3 (partnerships and collaboration), and 4 (expanding access to data through innovation) for Biodiversity Science.

5 Potential collaborations

A key component of this project is conducting integrative science to immediately inform decision making. As such, strategic collaboration among those with expertise in disciplines of ecology, data science, and human science will be of great benefit to this project. The inherently transdisciplinary nature of the SAS and especially the CSAS&L provides an ideal environment within which this project will thrive. I have identified potential collaborators with whose support the value of this project can only improve.

5.1 USGS RO #S31 research team members and proposed advisors

The expertise within this research team is impressive and the breadth conveys the degree to which transdisciplinary work is important to a successful project in CSAS&L. This project will benefit from advisement by Drs. Jenni and Wellman. Incorporating economic constraints (e.g., monitoring or management action budgetary restrictions) into the decision-making process is critical to exploring management options. Advisement from Dr. Bagstad will be useful in the tool designing phase to code a program which can easily incorporate financial decisions.

5.2 Intra-agency scientists and research groups

The diversity of scientific backgrounds of research scientists in the USGS is amazing. Potential USGS collaborators whose expertise would benefit this project include but is certainly not limited to, Andy Royle (population modelling), Mevin Hooten (hierarchical/Bayesian modelling), Jenn Malpass (ornithology, bird banding), Kristin Byrd (decision support), and Michael Runge (SDM). The Interagency Collaborate for Environmental Modelling and Monitoring meetings (ICEMM) will also be of great value to this project and my understanding of agency and interagency collaborations.

5.3 Community for Data Integration (CDI)

The technical aspects of this project, including software development and process documentation (especially 6.3.1) will benefit from regular interaction with the Community for Data Integration.

5.4 Working groups

Synthesis and integration of disparate data to inform a single model (Integrated Modelling) is aided by strong communications among the stakeholders, the Analyst Team and the subject experts. A formally-supported working group may expedite this communication, and birth new ideas for achieving aims beyond the scope of this project. I am aware of the potential for the Powell Center to facilitate in-person meetings of collaborators where funding may not otherwise be available. If appropriate, I will coordinate proposals for synthesis working groups (e.g., USGS Powell Center, NimBios) to explore further the challenges and opportunities of IPMs [10, 18–21]).

6 Expected products

The overarching aim of this project is to develop high-value Decision Support Tools (DST) for decision makers and placed-based within an Integrated Modelling framework. I proposed to create a DST that will be of high value to end-users, but with a broader goal of promoting the awareness and use of USGS data products and tools among the research and management communities. This project is designed to incorporate information (data) largely collected and curated by the USGS into advanced population modelling techniques, which can be used to directly inform on-the-ground management and decision makers.

Given the plethora of data and information available to applied ecologists and ecological management, it becomes increasingly important for the USGS to create awareness about their data products. This project promotes the awareness of USGS data projects while achieving our aims through both scientific and agency publications (6.1, 6.2).

This project will operate under an open science framework while adhering to both the USGS best practices for open source efforts [22] and Fundamental Science Practices. All efforts will be made to create publicly-available and easily accessible and executable end products.

6.1 Peer-reviewed scientific publications

I expect at least two high quality peer-reviewed publications from this project. The first will detail advances in the analytical and technical approach to building the IPM using a case study, where the target audience is similar to the readership of *Methods in Ecology and Evolution*. The second, an article focused on discussion of the utility of IPMs in the decision-making process, with a readership of applied ecologists and decision-makers. Close collaboration with decision-making teams will be required for a successful translation of the latter article to our intended audience.

6.2 Internal reports and outreach materials

In addition to regular internal reports required of the Mendenhall Fellow, I will generate a companion report that will be stored on the GitHub repository associated with these project files. Written for a general audience, this report will accompany the transparent process documentation of building our IPMs and DSTs. If possible, I will present our methodological results and DST end products at relevant meetings or conferences (e.g., The Wildlife Society annual meetings).

6.3 Decision support tools

I expect creating the decision support tools that are accessible and useful to the decision makers and practitioners to be the most challenging aspect of this project. Although generating the statistical programs (6.3.1) and web-based, interactive inference tools (6.3.2) are technically feasible, gaining feedback from our potential end-users will be vital to generating a *high-value* DST.

6.3.1 Statistical programming code and process documentation using GitHub repositories

Version control is essential for rapid and transparent development of this and other modelling efforts. Code and process documentation for building and analyzing Integrated Population Models will be made publicly available through a GitHub repository. Because the IPMs require at least a basic understanding of how to import and manipulate data, the end-user ranges from the intermediate to advanced statistical programmer. However, an R Package will be created to lower the threshold of understanding of coding, but long-term maintenance of the package is suggested. GitHub repositories are also an outlet for soliciting and documenting feedback on both the code and for building the models.

6.3.2 Web-based, interactive inference tools

Ideally, a web-based DST will be available for building custom IPMs, in addition to the reproducible scientific programming code (see 6.2.1). If possible, I will work closely with software and web developers within SAS to design a web-based interface for building and analyzing IPMs, and place-based data.

7 References

1. Roughgarden J (1997) Production Functions from Ecological Populations: A Survey with Emphasis on Spatially Implicit Models. In: Spatial ecology: the role of space in population dynamics and interspecific interactions. Princeton University Press
2. Boyd D, Crawford K (2012) Critical Questions for Big Data. Information, Communication & Society 15:662–679. <https://doi.org/10.1080/1369118X.2012.678878>
3. La Sorte FA, Lepczyk CA, Burnett JL, et al (2018) Opportunities and challenges for big data ornithology. The Condor 120:414–426

4. Hampton SE, Strasser CA, Tewksbury JJ, et al (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment* 11:156–162. <https://doi.org/10.1890/120103>
5. Jenni KE, Goldhaber MB, Betancourt JL, et al (2017) Grand challenges for integrated USGS science—A workshop report. U.S. Geological Survey, Reston, VA
6. Schaub M, Gimenez O, Sierro A, Arlettaz R (2007) Use of Integrated Modelling to Enhance Estimates of Population Dynamics Obtained from Limited Data. *Conservation Biology* 21:945–955. <https://doi.org/10.1111/j.1523-1739.2007.00743.x>
7. Ahrestani FS, Saracco JF, Sauer JR, et al (2017) An integrated population model for bird monitoring in North America. *Ecological Applications* 27:916–924
8. Jenni KE, Pindilli E, Bernknopf R, et al (2018) Multi-Resource Analysis—Methodology and synthesis. U.S. Geological Survey, Reston, VA
9. Schaub M, Abadi F (2011) Integrated population models: a novel analysis framework for deeper insights into population dynamics. *J Ornithol* 152:227–237. <https://doi.org/10.1007/s10336-010-0632-7>
10. Arnold TW, Clark RG, Koons DN, Schaub M (2018) Integrated population models facilitate ecological understanding and improved management decisions. *The Journal of Wildlife Management* 82:266–274
11. Zipkin EF, Saunders SP (2018) Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation* 217:240–250. <https://doi.org/10.1016/j.biocon.2017.10.017>
12. Sauer JR, Hines JE, Fallon J (1966) The North American breeding bird survey. Results and analysis 2003:
13. Besbeas P, Freeman SN, Morgan BJ, Catchpole EA (2002) Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* 58:540–547
14. Freeman SN, Crick HQP (2003) The decline of the Spotted Flycatcher *Muscicapa striata* in the UK: an integrated population model. *Ibis* 145:400–412. <https://doi.org/10.1046/j.1474-919X.2003.00177.x>
15. Coates PS, Prochazka BG, Ricca MA, et al (2018) The relative importance of intrinsic and extrinsic drivers to population growth vary among local populations of Greater Sage-Grouse: An integrated population modelling approach. *tauK* 135:240–262. <https://doi.org/10.1642/AUK-17-137.1>
16. Ross BE, Haukos DA, Hagen CA, Pitman J (2018) Combining multiple sources of data to inform conservation of Lesser Prairie-Chicken populations. *tauK* 135:228–240. <https://doi.org/10.1642/AUK-17-113.1>
17. Chandler RB, Clark JD (2014) Spatially explicit integrated population models. *Methods in Ecology and Evolution* 5:1351–1360
18. Hostetler JA, Sillett TS, Marra PP (2015) Full-annual-cycle population models for migratory birds. *The Auk* 132:433–449. <https://doi.org/10.1642/AUK-14-211.1>
19. Ahrestani FS, Saracco JF, Sauer JR, et al (2017) An integrated population model for bird monitoring in North America. *Ecological Applications* 27:916–924
20. Fink D, Auer T, Ruiz-Gutierrez V, et al (2018) Modelling Avian Full Annual Cycle Distribution and Population Trends with Citizen Science Data. *bioRxiv* 251868
21. Zipkin EF, Saunders SP (2018) Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation* 217:240–250
22. U.S. Geological Survey (2018) Documents best practices for software development within the U.S. Geological Survey: [usgs/best-practices](https://github.com/usgs/best-practices). <https://github.com/usgs/best-practices>. Accessed 7 Jan 2019