Proceedings of the Fourth International
Partners in Flight Conference: Tundra to Tropics
396–411

DATA FOR DEVELOPING SPATIAL MODELS: CRITERIA FOR EFFECTIVE CONSERVATION

Neal D. Niemuth,[1] Michael E. Estey, and Ronald E. Reynolds

Habitat and Population Evaluation Team, U.S. Fish and Wildlife Service, 3425 Miriam Avenue, Bismarck, North Dakota 58501, USA

Abstract. A major focus of North American bird conservation programs is the development of spatial planning tools to guide local and regional conservation efforts. A variety of databases is available to aid the development of spatial planning tools for bird conservation. These data sets are attractive because they are available for most of North America, are Internet accessible, and often are available in GIS format, seemingly making them ideal for spatial applications. However, the characteristics and quality of data available for bird conservation vary, which can greatly affect results of planning tools developed with the data. To be useful for on-the-ground conservation, spatial planning tools require data that are spatially and thematically accurate, fine-grained, have minimal bias, and were collected at the proper time. Bird data collected with limited planning or purpose suffer from limitations caused by variable sampling effort, inconsistent sampling protocol, lack of a sampling framework, small sample size, poor timing, inclusion of opportunistic observations, and a tendency to report unusual observations. Similarly, classification accuracy, spatial and thematic resolution, timing, and consistency of landcover data vary widely. Data should not be used uncritically. How data are collected, processed, and disseminated will have a great effect on their use, the quality of resultant spatial tools, and the ability of these tools to provide useful guidance for decision making. No data or methods for developing useful spatial planning tools are perfect, but some are preferable to others. We present factors to consider when evaluating data for use in spatial tools, provide examples of consequences of not considering those factors, and provide recommendations for use of spatial data.

Key Words: conservation design, conservation planning, data quality, GIS, geographic information system, spatial model, spatially explicit habitat model.

DATOS PARA ELABORAR MODELOS ESPACIALES: CRITERIOS PARA LA CONSERVACIÓN EFICAZ

Resumen. Un foco de central atención dentro de los programas de conservación de aves en Norteamérica, está en el desarrollo de herramientas de planificación espacial que sirvan de guía a los esfuerzos de conservación locales y regionales. Una diversidad de bases de datos está ya disponible para asistir en el desarrollo de estas herramientas para la conservación de aves. Estos sets de datos resultan en sumo sugestivos pues están disponibles en la mayor parte de América del Norte, son accesibles a través de la Internet y a menudo se encuentran en formato SIG, lo cual, combinado, les hace ostensiblemente ideales para aplicaciones espaciales. Sin embargo, las características y la calidad de los datos citados para la conservación de las aves varían, lo cual puede afectar profundamente los resultados de las herramientas de planificación, elaboradas a partir de ellos. Para ser de utilidad para la conservación en el terreno, las herramientas de planificación espacial requieren de datos que sean espacial y temáticamente precisos, minuciosos, con el mínimo de parcialidad y recogidos en los momentos adecuados. Los datos sobre aves recogidos con planificación o propósito limitado, sufren de limitaciones causadas por esfuerzos de muestreo variable, incoherente protocolo de muestreo, falta de un marco de muestreo, pequeño tamaño de las muestras, pobre coordinación, inclusión de observaciones oportunistas y tendencia a reportar observaciones inusuales. Asimismo varían grandemente los datos relativos a la precisión de la clasificación, la resolución espacial y temática, la coordinación y la consistencia de la cobertura terrestre. Los datos no deben ser utilizados sin sentido crítico. El modo en que se recojan, procesen y divulguen los datos, tendrá un amplio efecto sobre su propio uso, la calidad de las herramientas espaciales resultantes a partir de ellos y la capacidad de estas herramientas de proporcionar una guía útil en la toma de decisiones. No existen datos o métodos perfectos para desarrollar herramientas útiles de planificación espacial, pero algunos sí son preferibles a otros. En este trabajo presentamos factores a considerar a la hora de evaluar datos para su uso en herramientas espaciales, proveemos ejemplos de las consecuencias del no considerar esos factores y ofrecemos recomendaciones para el uso de datos espaciales.

[1] E-mail: Neal_Niemuth@fws.gov

INTRODUCTION

Concern over decreasing bird populations has stimulated a variety of bird conservation plans, many of which (e.g., North American Waterfowl Management Plan, Partners in Flight, The Nature Conservancy's Migratory Bird Program) explicitly promote a landscape approach to bird conservation. The increasing awareness of the importance of landscape composition to avian ecology and conservation, in conjunction with the recent upsurge in availability of spatial analysis software and data, has led to increased development and application of spatially explicit models to direct conservation actions (Wiens 2002). These models, often referred to as spatial planning, conservation design, or conservation assessment tools, provide a habitat-based context for conservation and are used for a variety of purposes, including identification of habitat and lands for protection, prioritization of funding and programs, and identification of opportunities for restoration.

Developing these tools requires large amounts of data, particularly when statistical analyses are used to characterize biological relationships. Data used for the development of spatial models should be thematically and spatially accurate for the desired use, spatially balanced, complete, and representative of the population of interest to ensure quality spatial tools. This applies to response (bird) and predictor (landscape) data used in model development. How data are collected, processed, and disseminated will have a great effect on their use, the quality of resultant spatial tools, and the ability of these tools to provide useful guidance for decision making.

In recent years, the number of datasets available for bird conservation planning has increased dramatically. These datasets are attractive for several reasons, as many use volunteers, are widely available, and are accessible from the Internet. Some of the data are available in GIS format, seemingly making them ideal for spatial applications. However, increased availability of spatial data for bird conservation planning presents potential problems as well as opportunity. The opportunity lies in the ability to use these data to develop models and tools that can better guide conservation actions. This differs from a more traditional use of bird survey data where populations are monitored and, after some time interval, identified population declines alert the conservation community to a problem, at which point remediation efforts begin. Caution must be used with readily available datasets as data that were collected for other purposes might provide poor or misleading guidance for decision makers when used to develop spatial planning tools.

Characteristics of data and concerns about data quality are hardly new topics, but the increased availability of spatial data over the Internet and increased emphasis on using these data to develop spatial models warrant new cautions. Just because data are available, or are even the "best available," does not mean the data are pertinent to the issues at hand or are suitable for use (Krebs 1989, Anderson 2001). Data must be appropriate for the intended purpose to ensure that spatial models are biologically sound and genuinely useful for guiding conservation. Unfortunately, data may be biased, inaccurate, or otherwise poorly suited for developing spatial models. Consequently, naive use of these data in spatial models may result in misdirected efforts, wasted resources, or unintended consequences. Data used to develop spatial tools should be rigorously evaluated for completeness and accuracy, and in all cases, the biology of species as it relates to the purpose of the planning tool and the planned conservation treatment must be kept in mind.

> *GISs encourage the user to do things that are often not justified by the nature of the data involved.*
> Goodchild and Gopal (1989)

We discuss and provide examples of data characteristics and how they can influence spatial tools. Problems that we identify relate to methods of data collection, as well as processes used to manipulate, display, transfer, and use data. Some of the latter border on modeling, which can also greatly influence the development and application of spatial tools, but is beyond the scope of this paper (for more information on spatial modeling see Scott et

> *Perhaps research and management biologists fail to realize how untrustworthy inferences from subjective, convenience samples can be…. Numbers from such surveys will not provide a basis for reliable knowledge and will represent only wasted resources.*
> Anderson (2001)

al. 2002, Woodbury 2003, Beauvais et al. 2006, Millspaugh and Thompson 2008). We offer solutions to help identify and address problems with data used in the development of spatial models for conservation planning. The issues we identify are not unique to the examples we provide but are inherent to most, if not all, data sets to varying degrees. Therefore, all data, regardless of their source, should be critically assessed prior to their use as to how well they meet the needs of the people and programs that will be using them, as well as how well they meet the assumptions and limitations of the methods used to analyze the data.

SPATIAL ACCURACY

A fundamental requirement of spatial data is that it be spatially accurate. This applies to both local and coarse scales. Obviously, point locations of species observations must be accurate if observations are to be linked to landcover data associated with that point to identify habitat associations and create spatial models at local or regional scales. But spatial accuracy is also important for very general, coarse-grained information such as species distribution maps. The distribution of bird species may be poorly known, and actual distribution can vary greatly among years. Field guides typically show generalized species range maps that are inherently imprecise, as a 2.5 cm-wide map of North America printed in a field guide is portrayed at scale of approximately 1:200 000 000. But the physical act of digitizing a range map creates exact coordinates for polygon boundaries, which may be interpreted as a level of precision that was not present in the data. In addition, maps digitized on-screen may introduce error and can misplace populations.

We compared range maps for Greater Prairie Chicken (*Tympanuchus cupido*) that we downloaded from the Internet (www.NatureServe.org) to data digitized from maps developed by Greater Prairie Chicken researchers and managers in Westemeier and Gough (1999) and associated references. The Greater Prairie Chicken is an area-sensitive grassland bird that has been extirpated from most of its former range, is in danger of extirpation in much of its remaining range, and is a species of conservation concern on a variety of prioritization lists (Johnsgard 1973, Schroeder and Robb 1993, Rich et al. 2004). Assuming that the map developed from Westemeier and Gough (1999) more accurately represents the range of the Greater Prairie Chicken, the map downloaded from the Internet missed major portions of the Greater Prairie Chicken's range in Kansas, Oklahoma, and Missouri, as well as smaller populations in Illinois, Iowa, Wisconsin, Minnesota, Missouri, South Dakota, and North Dakota (Fig. 1). In addition, the downloaded dataset included large areas that are not inhabited by Greater Prairie Chickens, and misplaced a population in Minnesota (Fig. 1). Data such as these are not suitable for conservation planning, even at a coarse scale. Regardless of which is correct, it is obvious that the results– and therefore the value– of spatial tools can be strongly influenced by the choice of data set.

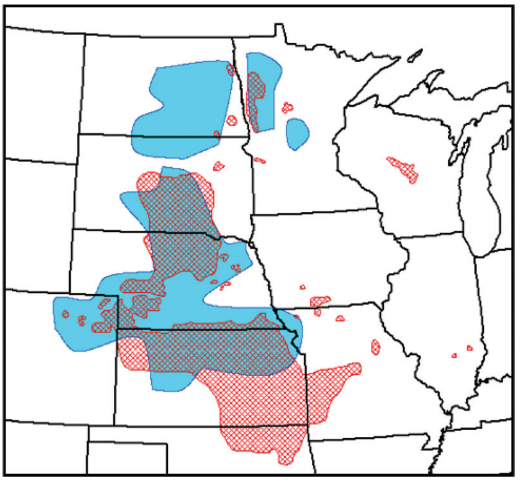


FIGURE 1. Apparent distribution of species varies among data sources, as range of Greater Prairie Chicken in the United States according to NatureServe (blue) differs from range adapted from Westemeier and Gough (1999; red) and associated references. NatureServe data provided in collaboration with Robert Ridgely, James Zook, The Nature Conservancy Migratory Bird Program, Conservation International Center for Applied Biodiversity Science, World Wildlife Fund US, and Environment Canada WILDSPACE.

Providers of data such as species distribution maps typically provide caveats to the effect that the data are coarse grained and are for general planning purposes only, and that data quality should be verified or other data acquired for specific planning. However, the justification of using such data only for coarse-grained analyses begs the question as to the value of such coarse-grained data, as planning must be at least somewhat specific to have any value for on-the-ground treatments such as management or acquisition. Even at extremely coarse scales, though, data should not be used uncritically, as model predictions can change greatly when multiple range maps with spatial errors are overlaid (Dean et al. 1997).

SPATIAL BALANCE AND SAMPLING BIAS

Spatial data should be spatially balanced and unbiased so the data are representative of the population of interest across the entire study area. These characteristics will be greatly influenced by the sampling framework and protocol used to collect the data, as well as the species being considered. Data from breeding bird atlases illustrate this issue well, but again, the issues we pose are present to some degree in many datasets. Some atlases attempt to address problems associated with unequal sampling effort by dividing a region into grid blocks and hiring biologists to sample blocks that volunteers have not surveyed. However, the issue remains that sampling effort varies among blocks, which, if not considered in analysis, will affect inferences made from the data and their value for conservation planning. This may be more problematic with some species than others and can vary depending on detectability and distribution of the species being considered. Again, data should always be assessed in terms of biology and intended use rather than simply accepted or rejected.

For example, we considered the distribution of Ruffed Grouse (*Bonasa umbellus*), Ovenbird (*Seiurus aurocapillus*), and Barred Owl (*Strix varia*) in Wisconsin as documented by the Wisconsin Breeding Bird Atlas (Cutright et al. 2006) relative to landcover as classified by the Wiscland landcover database (Gurda 1994; Fig. 2A). Wiscland data are based on satellite imagery collected primarily in 1992 and follow a three-level hierarchical classification scheme (metadata including accuracy assessment are available at http://dnr.wi.gov/maps/gis/datalandcover.html). According to Wiscland, the northern third of the state is heavily forested, the southeastern third is predominantly agricultural and urban, and the middle third is a mixture of woodlands, wetlands, and agriculture. Given the distribution of forest cover, one would expect most observations of these three forest species to occur in the northern third of the state, where forest cover is dominant. This is indeed the case for Ruffed Grouse and Ovenbird, and observations of these species during the Wisconsin Breeding Bird Atlas closely reflect the distribution of forest cover throughout the state (Figs. 2B and 2C). However, atlas data show more confirmed breeding Barred Owls in the central and southern parts of the state, which have relatively little forest, than the northern part of the state (Fig. 2D). There are possible biological explanations for these patterns: Barred Owls may be less forest dependent than Ovenbird or Ruffed Grouse, or Barred Owls may be more likely to be found in fragmented forests than Ovenbird or Ruffed Grouse. Or it could simply be that Ruffed Grouse and Ovenbirds are diurnal, occur at relatively high densities, and are easily detected during the time periods when most birders are active, whereas Barred Owls are nocturnal, occur at relatively low densities, and are best detected at night when fewer birders are active. Consequently, apparent Barred Owl distribution likely reflects survey effort as well as habitat, as northern Wisconsin has few people compared to central and southern Wisconsin. In short, known presence of a species at a site may be a reflection of survey effort rather than habitat quality. The relationship between survey effort and detected species richness has been well documented; *r* values between species richness and sampling effort include 0.57 for Pennsylvania birds in a coordinated breeding bird atlas (Brauning 1992) and 0.98 for plants in a biodiversity database (Hortal et al. 2007).

Natural Heritage Inventory data are frequently suggested as data sources for planning and evaluation tools in bird conservation efforts. These data are attractive as they are available from established programs, are available for most of the North American continent, and are often accessible from the Internet. In many cases, the data come with extensive metadata. However, Natural Heritage Inventory data may not be appropriate in some cases or available for some species of interest.

First, given limitations of funding and staff, some natural heritage databases must limit the number of species for which they include data, with priority justly given to threatened and endangered species. Second, many of the data were collected and reported opportunistically (Beauvais et al. 2006, Stein et al. 2008) and inferences made from the data will be influenced by variable sampling effort, inconsistent sampling protocol, inclusion of opportunistic observations, and a tendency to report unusual observations. Such data are not reliable for making inferences about wildlife populations (Anderson 2001) and models, program assessments, and conservation

> *Considerable quantities of ecological survey data have accumulated over the last few decades, but there is little information on its reliability. Indeed, ecologists in general have been slow to address questions of data quality and observer error.*
>
> Cherrill and McClean (1995)

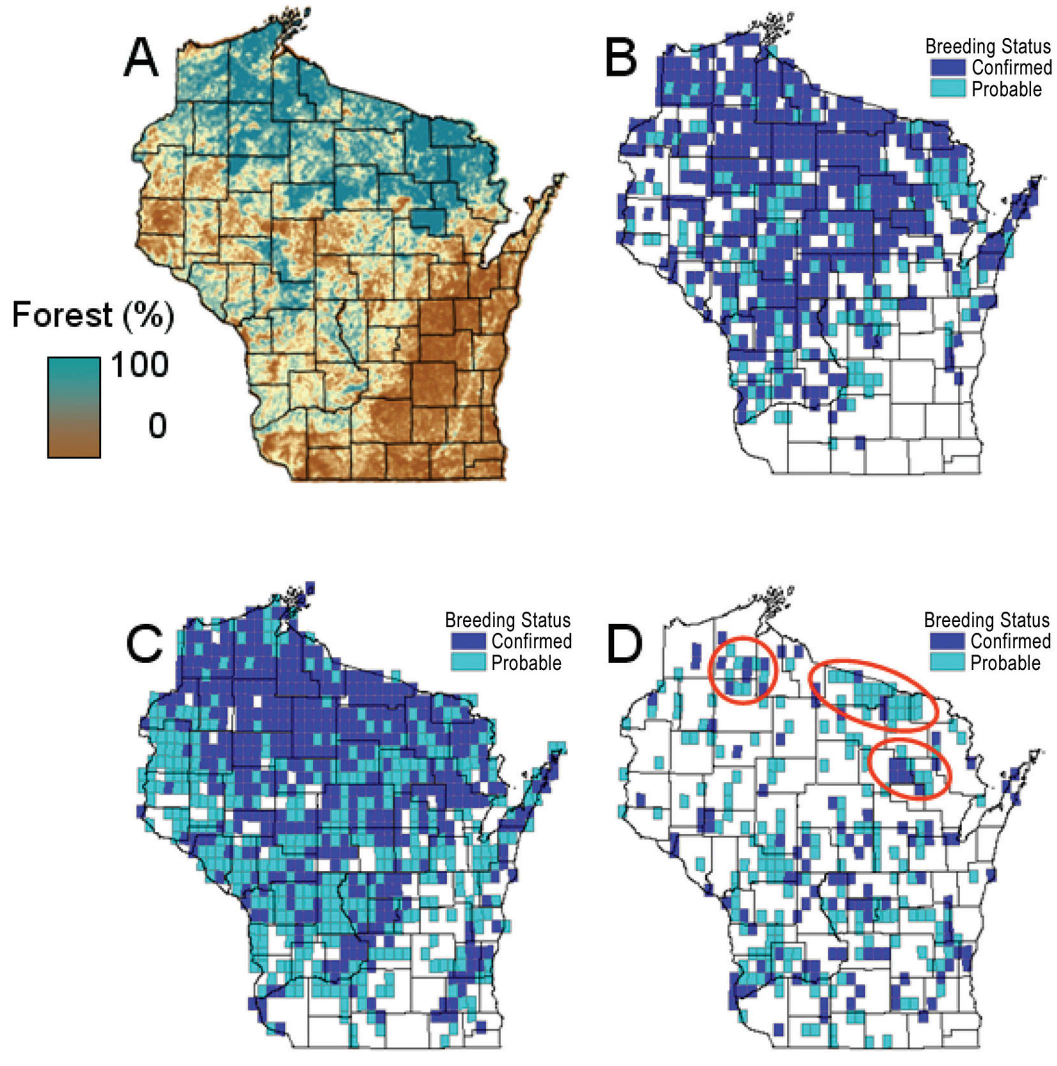


FIGURE 2. Detection and apparent distribution of forest birds vary among species. (A) Percent of the Wisconsin landscape classified as forest within an 800-m radius moving window. (B) Breeding status of Ovenbird, (C) Ruffed Grouse, and (D) Barred Owl in Wisconsin as determined by breeding bird atlas. Red circle and ovals in (D) indicate areas in northern Wisconsin where owl research or nocturnal surveys took place.

efforts based on anecdotal occurrence data may be flawed (McKelvey et al. 2008). Consequently, opportunistic observations should be carefully filtered for negative data, element identity, mapping precision, historical records, season of occurrence, and extra-limital records (Beauvais et al. 2006).

We viewed data from the North Dakota Natural Heritage Inventory database and found 109 records for the endangered Least Tern (*Sterna antillarum*), widely distributed throughout the limited range of this species in the state (Fig. 3A). In this case, the large number of records suggests that the data may be the result of a coordinated survey effort, and that the data warrant further consideration. However, the same database had only one observation for the much more abundant and widespread Black Tern (*Chlidonias niger*), reported from a predominantly wooded state park outside the areas where Black Terns are most likely to be found in North Dakota (Fig. 3B). Clearly, the single observation indicates that these data are insufficient for development of spatial tools for Black Terns, but even if there were more observations, the unusual

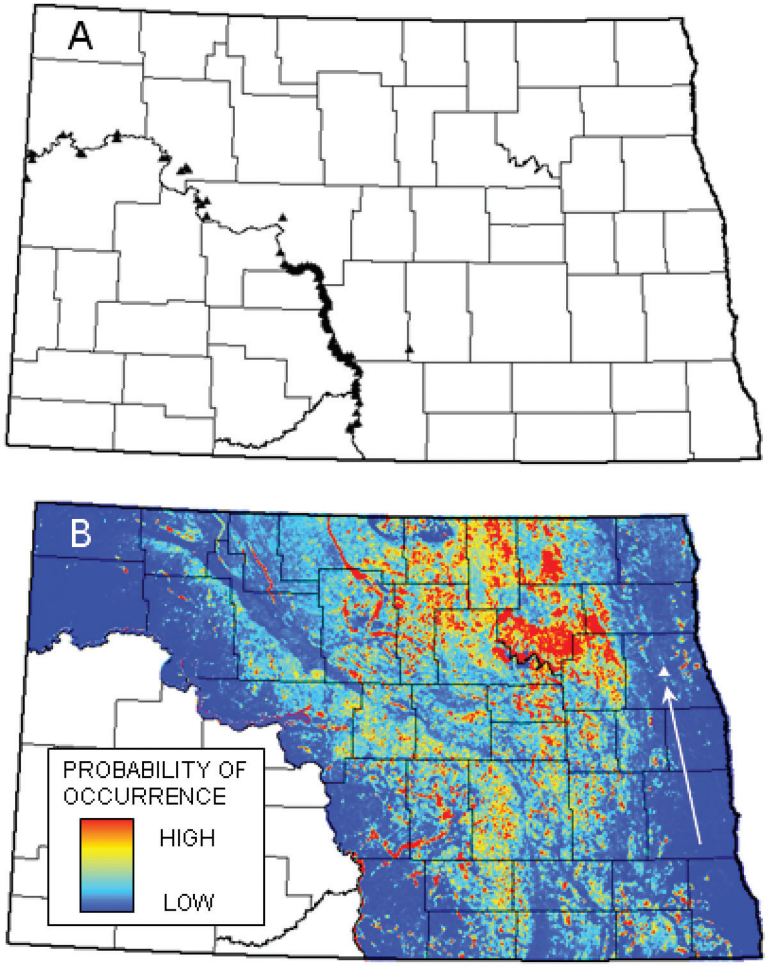


FIGURE 3. (A) Records of the rare Least Tern (black triangles) in North Dakota from North Dakota Natural Heritage Inventory data greatly outnumber (B) record of Black Tern (white triangle at tip of white arrow). Probability of occurrence based on associations between Black Terns and landscape characteristics (including wetlands) following methodology outlined in Niemuth et al. (2008).

location of this sighting would warrant caution about how well it represents Black Terns in North Dakota.

COMPLETENESS AND RESOLUTION OF RESPONSE DATA

A problem with many species presence datasets is the lack of absence data (areas where birds were not present or were not detected during surveys), which complicates analysis and limits inferences that can be made from the data. For example, species presence data are sometimes used to identify all cover types in which a species was found, and this model of habitat use is then applied across the landscape. The usefulness of a spatial tool can be greatly increased by inclusion of data on species absence as well as presence, as this enables determination of what

All the fancy algorithms, statistical gymnastics, and mapping tricks in the world can't help poor input data. … 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.
Beauvais et al. (2006)

habitats birds used in relation to what was not used. Although the same species presence data may be used, the analyses differ in that the first simply documents use whereas the second determines selection. This difference is not trivial, as two very different products can result from the exact same data, with tremendous implications for conservation planning. The second approach requires better data, as well as additional thought and processing, but provides far stronger inferences and bases for conservation.

An example of this is a recent conservation plan developed within the U.S. Fish and Wildlife Service that prioritized landscapes for conservation in the northern Great Plains based in part on presence models for 13 species of grassland birds. According to model descriptions, 10 (77%) of the 13 species had been found in dry agriculture or irrigated agriculture crop fields, or both, as well as grassland. Consequently, spatial models for these 10 species identified agricultural lands and grasslands where the species could be expected to be found. This was reasonable, as the original purpose of the models was to show where birds might occur with no regard to density or demographics. However, the conservation plan used these models to prioritize landscapes, assigning equal points to agricultural lands and grasslands for conservation of these 10 species. Grassland birds have experienced the steepest, most consistent, and widespread population declines of any bird group in North America (Knopf 1996) primarily because of conversion of grasslands to agriculture. But according to the conservation plan in question, lands in grass would receive the same score after being plowed up and converted to crop fields. This is biologically unsound, and illustrates how the value of data is determined in part by the question being asked and how the data are used.

A variety of techniques exist for analyzing presence-only data, but many are sensitive to sample size, extent of species distribution, and degree of specialization of species being considered (Elith et al. 2006, Hernandez et al. 2006). Correctly analyzed and presented, data collected from well-designed presence-absence (Sauer 2008), occupancy (MacKenzie et al. 2006), or density (Buckland et al. 2004) surveys will provide better biological information, overcome shortcomings associated with presence-only data, calibrate estimated frequency of occurrence, and reduce uncertainties and the false sense of precision associated with some habitat models and maps (Elith et al. 2002, Guisan et al. 2005, Elith et al. 2006).

SPATIAL RESOLUTION OF DATA AND PLANNING TOOLS

Proper resolution of data and analysis are critical to understanding biological relationships, as well as the development and implementation of spatial planning tools. Changes in extent and grain of study areas and data can significantly alter perceptions of landscape composition and habitat selection (Wiens 1989), which will in turn affect development and interpretation of spatial tools, as well as subsequent recommended conservation actions.

Many data are available or used at a very coarse resolution (i.e., EPA hexagons, 7.5′ quadrangles, civil townships, or watersheds). Data with extremely coarse resolution may provide some general snapshot of national conditions, but will likely provide no new information and are susceptible to problems of scaling. Because of these issues and other limitations, coarse-grained data are rarely appropriate for planning specific conservation actions. Even if fine-grained models are being considered, spatial errors associated with coarse-grained biological response data may prevent development of meaningful models, as actual locations and landcover types used by birds may not be known.

> *Data may be useless for several reasons. They may be unreliable or unrepeatable. They may be perfectly reliable and accurate but irrelevant to the problem at hand. They may be reliable, accurate, and terribly relevant but not collected at the right season of the year. Or the experimental design may be so hopeless that a statistical analysis is not possible.*
> Krebs (1989)

Breeding bird atlases can provide useful information on the broad-scale distribution of species, but the coarse resolution of many atlases (e.g., portions of civil townships or USGS 7.5′ topographic quadrangles) and lack of habitat linkages precludes planning at the resolution at which conservation actions take place. More recent efforts have linked bird observations and habitat (e.g., Great Basin Bird Observatory 2005), which provides a stronger biological foundation for conservation.

Even if habitat characteristics were recorded where species were detected, the resolution at which data were processed and models

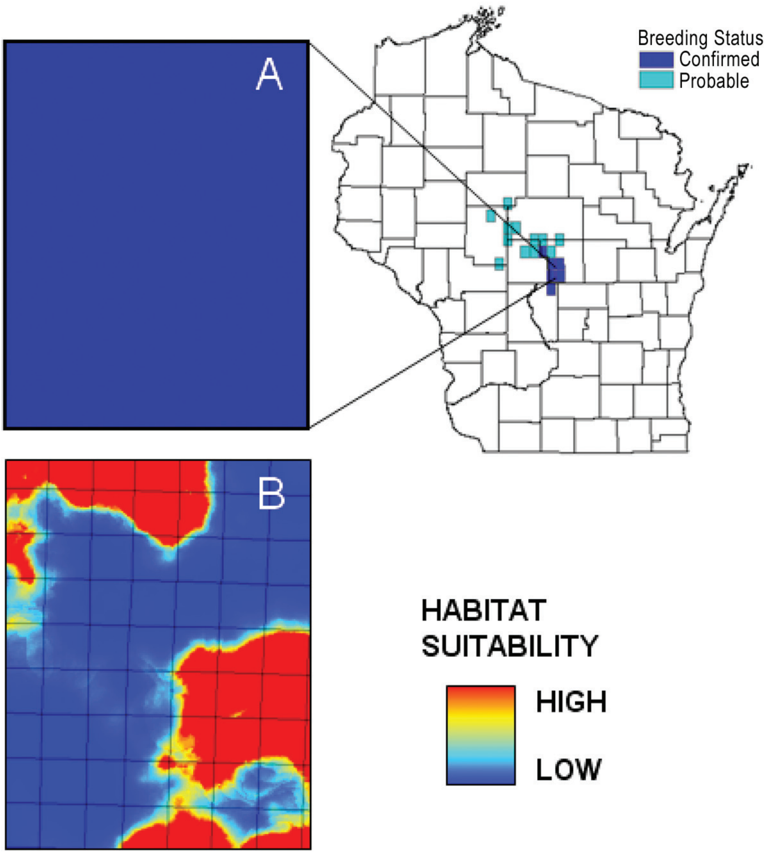


FIGURE 4. Resolution of and information in spatial representations for conservation planning vary among methods. (A) Breeding status of Greater Prairie Chicken in western Portage County, Wisconsin, as determined by breeding bird atlas has coarse spatial resolution and provides little information for local action. (B) Modeled habitat suitability for Greater Prairie Chickens in the same area (Niemuth 2003) shows a range of values at a much finer spatial scale, providing additional information that allows targeting of lands at scale usable by planners and managers. Gridlines in (B) indicate section lines at 1.6-km intervals.

developed determines the value of atlas projects for assessing conservation value or guiding conservation actions. For example, data from the Wisconsin Breeding Bird Atlas (Cutright et al. 2006) show blocks where Greater Prairie Chickens were recorded, but coarse-grained blocks were simply assigned a value indicating species status (Fig. 4A). Summary maps provide no indication of where the birds were within an atlas cell, their densities, or what types of habitat were present. A spatial model developed using landcover data and Greater Prairie Chicken lek locations in the same area (Fig. 4B) is more useful because it provides information at a finer resolution and articulates relationships between Greater Prairie Chickens and landscape characteristics. Because Greater Prairie Chickens were associated with the amount of grassland and wetland, proximity to other populations of Greater Prairie Chickens, and the absence of forest cover, planners and managers can assess the potential effect of conservation actions such as tree removal and grassland restoration on habitat for Greater Prairie Chickens.

Information loss can be especially problematic when large polygons (e.g., watersheds) are used to denote the presence of features recorded as points. Watersheds such as those identified by the United States Geological Survey's 8-digit hydrologic unit codes (HUCs; Fig. 5) are often used as an analytical unit when reporting occurrence of biological elements such as bird observations. Using watersheds as sampling units or to report species occurrence reduces the value of data as resolution is coarse, habitat heterogeneity within a watershed is masked, and

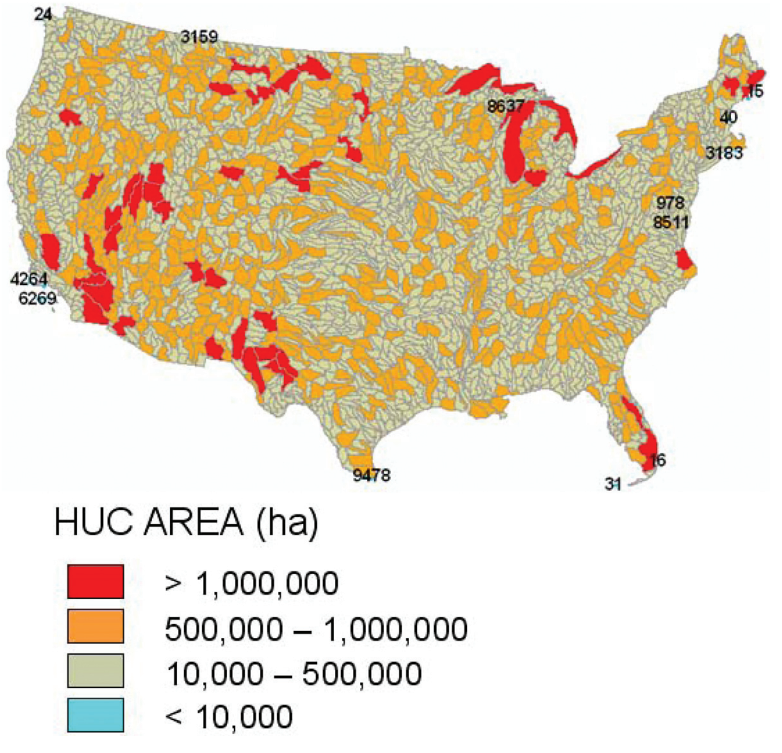
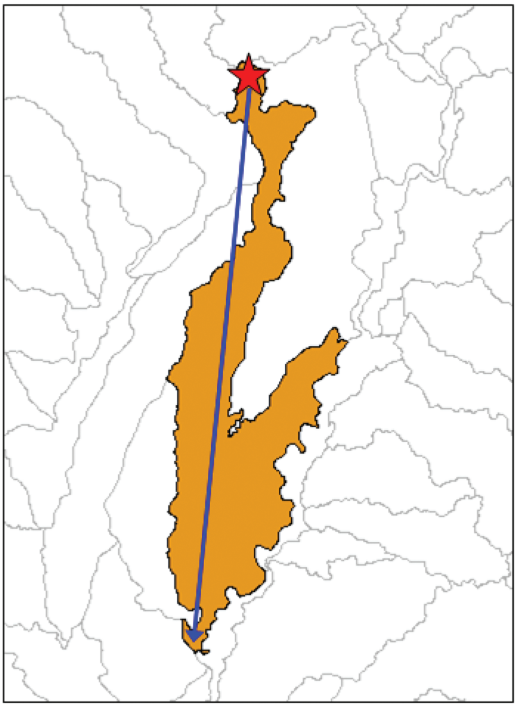


FIGURE 5. Size of watersheds identified by eight-digit Hydrologic Unit Codes (HUCs) in the conterminous United States varies dramatically. Digits indicate area (ha) of HUCs < 10 000 ha.



watershed boundaries may have little biological meaning relative to the conservation question being addressed.

For example, in an analysis conducted for a U.S. National Wildlife Refuge, element occurrences from a watershed that came within 2 km of the refuge boundary were ignored because they were in a different watershed, while occurrences within the watershed the refuge occupied were counted, even though portions of the watershed were 260 km from the refuge (Fig. 6). Data that are processed and reported in such a coarse manner will have little value for guiding conservation actions, and will likely be subject to scale-related errors of omission and commission.

In addition, different-sized and different-shaped watersheds are likely to introduce area effects into the data and subsequent models. The area of 8-digit HUCs in the conterminous

FIGURE 6. (*left*) Size of eight-digit Hydrologic Unit Code (HUCs) and location of national wildlife refuge (red star) in relation to HUC boundaries influenced how data were used to estimate diversity on the refuge. Distance from refuge to the far end of the HUC is 260 km.

United States ranges from <1 ha to >2.4 million ha, excluding the Great Lakes, which are larger (Fig. 5). When determining species presence or richness, sample units serve as the "net" by which species or threats are captured and recorded. Variably sized watersheds result in variable sampling effort and larger watersheds are more likely to have inflated estimates of species presence or richness due to passive sampling (Connor and McCoy 1979).

A common practice in spatial planning is to use large units such as watersheds for an initial, coarse-grained analysis to identify general priority areas and then use finer-grained information to more precisely identify priority areas within the coarse-grained priority zones. This process is problematic for two reasons.

First, an initial analysis that uses poor quality data and ignores issues of scale and variable sampling effort will do a poor job of identifying priority areas due to errors of omission and commission (Fig. 1) and thereby produce misleading results. Consequently, finer-grained, local analysis conducted within priority areas identified by the coarse-grained analysis will only prioritize a poorly selected portion of the landscape, disregarding areas with potential value for conservation action that could not be identified due to errors of omission in data used in the coarse-grained analysis.

Second, an additional, fine-grained analysis would provide more information if it were conducted across the complete area of interest, rather than a poorly selected subset of the area of interest. Hence, there is little reason to conduct the initial, coarse-grained analysis. The two-step approach would work if the only errors were errors of commission, but this would require a very inclusive approach to delineating species ranges and cataloging occurrences, which could verge on a comprehensive analysis of the entire area of interest. Performing one analysis using scales and data appropriate to the biology of the species and the program implementing conservation actions will provide one useful product, as opposed to two potentially flawed products. For all these reasons, resolution of fine-grained data should be maintained, and not degraded by data providers or users. One exception to this rule would be to protect rare species by only providing generalized, rather than high precision, locations, but this would only apply to data made available to the public, not data used for modeling.

Selection of appropriate scales for biological response and resolution of spatial models should be carefully considered when evaluating data, as the response to scale will differ among species (Wiens 1989). It is important to note that, after fine-grained analyses are complete, users can still display boundaries of coarse-grained units to place model output in the context of watersheds or political boundaries such as counties.

DATA INTEGRITY

Geographic information systems provide an excellent opportunity to unwittingly launder data of poor or unknown quality. This can be accomplished in a variety of ways. For example, poor quality data (e.g., species presence as in Fig. 1) or subjective values (e.g., degree of threat) can be entered into a database and then linked to geographic areas such as watersheds. These values can then be mathematically manipulated or combined with other data, some of which may be objective and unbiased. Resulting tables or maps may have an objective or quantitative aura that, on closer inspection, is found to be false. Similarly, subjective evaluations of important blocks of habitat can be digitized, thus transforming subjective "circles on a map" into "spatially explicit habitat goals developed using the best available information." Finally, areas identified and delineated as having high species richness through multiple overlays might not even have target species present (Fig. 1). Unfortunately, such errors are often unknown, as species identity is typically lost or not reported when the overlay process is used to create maps of species richness.

How a question is posed and results are presented can have a significant influence on how data are used, as well as the biological foundation of the resulting spatial tool and its value for spatial planning. As mentioned previously, overlaying is a core GIS manipulation and lends itself to using "species richness" or similar measures as a biological response. Diversity metrics often are inappropriate as a response variable in models for conservation planning (Conroy and Noon 1996, Villard et al. 1998, Goldstein 1999). Richness analyses are problematic for a variety of reasons, including poor consideration of species involved, differing life histories and habitat requirements, and issues of scale (Noss 1987, Rodda 1993, Conroy and Noon 1996, Dean et al. 1997). These problems are exacerbated when richness measures are estimated from poorly developed and untested models developed from coarse habitat associations. An important reason for not using overlays of atlas or heritage data is that such summaries lack the biological mechanisms or habitat linkages present in a modeling approach, and simply indicate whether or not a species has been recorded within a block. This provides little

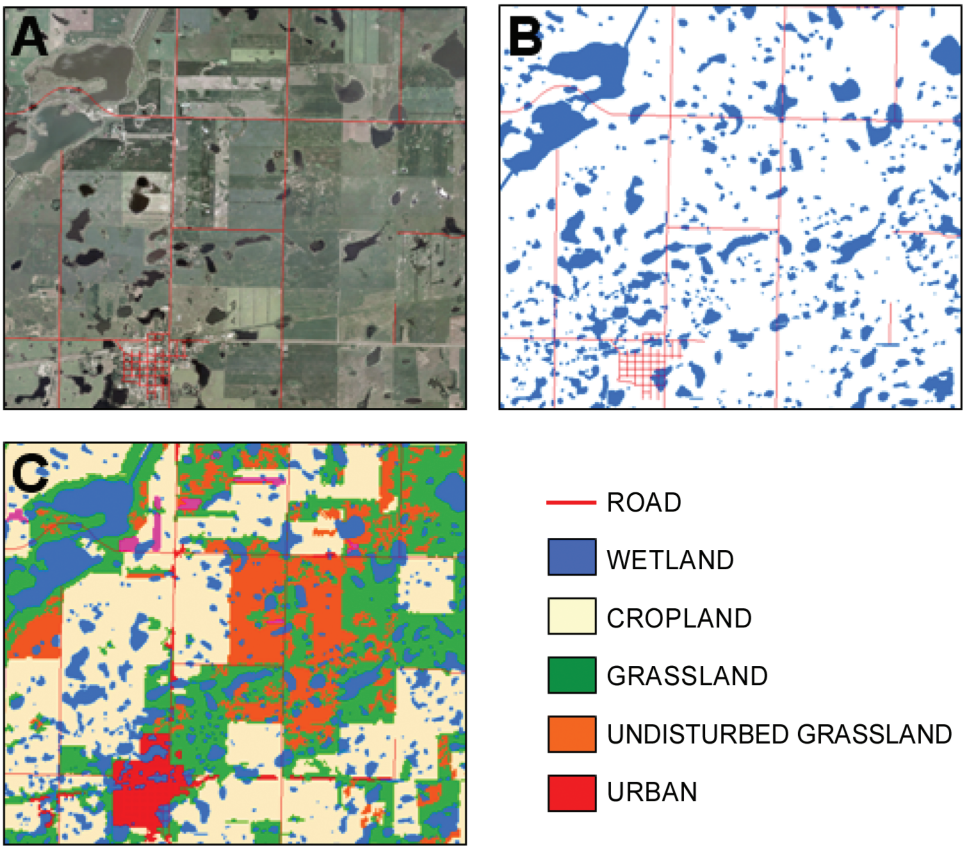


FIGURE 7. Amount and type of information that can be inferred from landcover data vary with source and application. (A) Aerial photograph of landscape in central Sheridan County, North Dakota, taken in 2003. (B) Wetland basins derived from 1979 NWI data for the same area. (C) Digital landcover data for the same area based on satellite imagery collected in 1992 with NWI wetland basins incorporated into layer. Non-urban roads follow section lines at 1.6-km intervals.

insight into the suitability of unsampled areas and precludes identification of specific sites for conservation or management action. Finally, the lack of mechanisms or habitat linkages does not allow identification of deficiencies in habitat (i.e., isolation, lack of key cover types) that can be addressed through appropriate management actions—which likely vary among species.

COMPLETENESS AND ACCURACY OF LANDCOVER DATA

The quality of any spatial planning tool also will be determined by the quality of landcover data used in developing and applying the habitat model. Landcover data can be acquired in several ways, some of which are inherently coarse-grained and subject to error. For example, the aerial photograph of a landscape in Fig. 7A contains considerable information that can be visually interpreted, but is not directly interpretable by a GIS. Fig. 7B shows wetlands in the landscape that were mapped, classified, and digitized by the National Wetlands Inventory (NWI), which provides information about wetlands and their location. Wetlands information can be combined with other digital data to help develop a more comprehensive landcover layer (Fig. 7C). However, wetlands were mapped and digitized as they appeared in 1979 during periods of optimal water conditions (Wilen and Bates 1995) and water levels and wetland area can vary considerably within and among years (Kantrud et al. 1989). Classified digital landcover data with 30-m resolution for the area (Fig. 7C) provides thematic classes not present in the aerial photograph (Fig. 7A), but the coarse thematic, temporal, and spatial resolution provide general information only.

Most digital landcover data that cover large spatial extents are based on satellite imagery that has been processed to separate signatures that can be associated with various landcover classes. However, classification of satellite imagery is subject to considerable error caused by a variety of factors including variation in land use and vegetation, shading, sensor variation, choice of landcover classes, timing, spatial error, and differences in phenology, soil types, and soil moisture (see Thapa and Bossler [1992], Green [1994], Lillesand and Kiefer [2000], Thogmartin et al. [2004]). Consequently, an assessment of classification accuracy should always be reported in any spatial analysis so users of spatial products have an idea of how reliable the data (and resultant tools) are. Unfortunately, many spatial analyses ignore this inevitable error and treat the data as though it were completely accurate, even though error in the landcover classification might considerably bias results of the spatial analysis.

Classification accuracy varies among datasets and even among classes within a dataset. For example, in a pixel-level assessment of National Landcover Data for six states in the northern Great Plains and Rocky Mountain Region (available at http://landcover.usgs.gov/accuracy/pdf/region8.pdf), user's accuracy for 20 landcover classes ranged from 1% to 79%, with an overall accuracy of 60%. Acceptable levels of error in any dataset will be determined by the goals and intended use of the conservation assessment. But clearly, the value of a spatial planning tool will be compromised by low accuracy of landcover data. Conservation planners using such data must be aware of its limitations and verify field conditions before implementing conservation actions. Also, the composition and configuration of landscapes will change over time; landcover data are snapshots that do not capture dynamics that may affect population responses. Limitations like this should be explicitly explained when presenting results of modeling efforts.

Differences in landcover classification are often apparent at boundaries such as state lines or the boundary between two adjacent satellite images. Users should be aware of landcover datasets that are clipped to irregular lines (e.g., Bird Conservation Region or watershed boundaries) or are "feathered' along edges. Clipping scenes to natural boundaries has a biological appeal, but these practices may simply be a strategy to reduce obvious changes in classification that are apparent along abrupt transition lines.

The effect of classification error on spatial models will depend in part on which landcover classes are incorrectly classified and which classes are confused. Typically, similar classes (e.g., mixed forest and coniferous forest) are more readily confused than dissimilar classes (e.g., plowed fields and coniferous forest). In some cases and when biologically appropriate, it may be useful to combine similar classes that are easily confused. There is a certain appeal to having many landcover classes (i.e., high thematic resolution), as multiple classes appear more biologically realistic in that they better represent the variety of landcover types present on the landscape. However, classification accuracy tends to decline as the number of classes increases, and the user may have to find an appropriate tradeoff between thematic resolution and classification accuracy.

CONCLUSIONS AND SOLUTIONS

Given recent emphasis on landscape-level conservation, interest in using spatially explicit data is high. However, adoption and application of spatial data should be driven by need, purpose, and biology, not enthusiasm. Landcover data are often referred to as "remotely sensed" because the platform from which the data were recorded was not in contact with the area under consideration (Lillesand and Kiefer 2000). In a similar vein, bird data can also be considered remote as data anomalies that negatively affect analysis or biological nuances important to analysis and interpretation may be lost when data are used by people who are remote from the individuals and populations that were sampled.

Given the diversity of applications for spatial data and the variety of ways in which data characteristics can influence results of spatial tools, it is difficult to prescribe universal recommendations or quality thresholds for spatial data. An awareness of data characteristics and how they influence models is paramount, as is a precise and clearly articulated purpose for the spatial tool and a knowledge of how it will be implemented. Both providers and users of data used in bird conservation planning should investigate and assess all data as they relate to

> *Unfortunately, the recent rapid increase in the availability and use of GIS and other software has not been accompanied by a concomitant increase in the use of techniques for assessing uncertainty in spatial data and spatial models.*
>
> Woodbury (2003)

TABLE 1. Factors and characteristics to be considered when assessing data for appropriateness to purpose; spatial and thematic accuracy; resolution and integrity; and bias with regard to development of spatial tools. This list is not comprehensive and does not include specific prescriptions, but is intended to stimulate investigation and understanding of how data characteristics affect spatial tools used for conservation.

Appropriateness of data to purpose
Has a specific, clearly defined purpose that includes end use been identified for the spatial tool?
Have the process and data been reviewed and logic-checked by biologists and modelers familiar with the species, geographic area, and purpose of the tool?
Do the data appear to be representative of the population of interest?
Are data available across the entire region of interest or will multiple datasets with differing characteristics be used?
Are landcover data classes appropriate for the intended purpose?
Does timing of landcover data represent conditions when bird data were collected?
Spatial and Thematic Accuracy
Have locations of data points and lines been determined to be accurate?
Is classification accuracy of landcover data known?
Is classification accuracy of pertinent landcover classes sufficient for the intended purpose?
Data Resolution and Integrity
Have metadata been reviewed and data providers consulted regarding the specific application of the data?
Is spatial resolution of data appropriate for the intended conservation treatment?
Are landcover data sufficiently fine to represent features of interest?
Is the level of response (i.e., presence/absence, density) sufficient for the intended purpose of the planning tool?
Are data allocated to coarse units such as watersheds?
Is sample size sufficiently large to meet requirements of analysis and provide precise parameter estimates?
Bias
Did sampling effort vary across time or throughout the area of interest?
Does the dataset contain opportunistic observations?
Were data collection points determined using a statistically based sampling design?
Were data collected using a consistent and appropriate sampling protocol?
Was timing of data collection appropriate to the purpose?
Have the data and proposed analyses been reviewed by a statistician who is aware of the intended conservation treatment and end use of the spatial tool?
Have biases or shortcomings of these or similar data been identified by other researchers?

the intended purpose and final application of the spatial tool (Table 1). Users should also consult a statistician and spatial modeler, as well as data providers if additional information is needed. No data will meet all criteria, but users should identify biases and errors in biological and landcover data, and understand how these biases and errors might affect spatial tools developed with the data. For example, factors such as survey effort can be incorporated into models, but such information must be known to do so. Data may be promoted as being the "best available," but users must keep in mind that that does not guarantee that the data are useful for the intended application.

Users should work within limitations of data and use a transparent process that relates to species biology and programs being addressed. Black-box software or coarse-grained analyses allow minimal insight into and control over analyses, reducing the opportunity to assess and address data problems. Users should avoid such approaches or else determine that algorithms and approaches used by software packages are appropriate to the specific purpose of the spatial tool. Finally, users should report assumptions, potential biases, and issues of accuracy regarding the data they used. If an accuracy assessment is not included in landcover metadata or at the site where landcover data were obtained, it may be necessary to conduct an assessment for the area covered by the spatial tool (see Lillesand and Kiefer 2000). At a minimum, the absence of accuracy data and the resulting uncertainty in model output should be reported.

Bird data collected on systematic surveys are available that can be used for guiding conservation of many bird species. For example, data from the North American Breeding Bird Survey (Sauer et al. 2008) can be used in conjunction with digital landcover data to develop spatial models for a variety of species (Newbold and Eadie 2004, Niemuth et al. 2005, Thogmartin et al. 2006a), although these data and models also have their own limitations that must be acknowledged (Thogmartin et al. 2006b). If data sets of suitable quality simply do not exist for an identified species, need, and area, it may be necessary to collect appropriate data, either by establishing a new survey or modifying an existing survey. The expense that collecting additional data incurs should be offset by the

> *... it is an expensive impropriety to maintain that an untrustworthy estimate is better than none.*
>
> Delury (1954)

increased efficiency in conservation delivery that is expected to be gained from having a robust spatial tool to guide decisions. If a spatial tool is not expected to provide increased efficiency in conservation delivery, the need for developing that tool should be re-assessed. As an example, the value of data from many atlas projects could be increased by conducting standardized surveys at a statistically designed sample of sites, particularly if the surveys incorporated estimates of detectability (i.e., Buckland et al. 2004, MacKenzie et al. 2006). Data from these surveys would not suffer from variable sampling effort and could be used to develop habitat associations that could then be extrapolated to portions of the landscape not sampled. This would be especially valuable in sparsely populated states where only a small portion of the state is sampled during atlas projects. When quality bird data are lacking and cannot be acquired, conceptual models can provide useful guidance based on biological relationships (Clevenger et al. 2002, Niemuth et al. 2005), although assumptions of the model must be clearly identified. Whatever data and methods are used, conservation planners must bear in mind that bird populations and land cover are variable and that data sets represent past conditions that might not be pertinent to present decisions.

Final responsibility for spatial tools lies with the modelers who acquire and use data, but data providers should help ensure that their data are accurate and unbiased. Data providers should strive to only provide high-quality data with accompanying metadata, realizing that some users will not read the metadata or understand the implications of using data in a manner inconsistent with the way it was collected. In some cases, providers might want to withhold data with known pitfalls. Data providers should expect and plan for spatial applications of their data by providing data that are accurately georeferenced and not simply assigned to a coarse unit (i.e., county, watershed, township) in which the observation occurred. Finally, providers and users should have an in-depth understanding of the processes required for developing useful spatial tools, as well as how these tools are used for on-the-ground conservation.

ACKNOWLEDGMENTS

We thank K. Duttenhefner for providing data from the North Dakota Heritage program and B. Harriman for giving permission to use graphics from the Wisconsin breeding bird atlas. We greatly appreciate helpful comments on earlier drafts of this paper by L. A. Brandt, W. H. Busby, F. P. Howe, T. D. Rich, and F. R. Thompson. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

LITERATURE CITED

ANDERSON, D. R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29:1294–1297.

BEAUVAIS, G. P., D. A. KEINATH, P. HERNANDEZ, L. MASTER, AND R. THURSTON. 2006. 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. Wyoming Natural Diversity Database, University of Wyoming, Laramie. [Online.] <http://www.natureserve.org/prodServices/pdf/EDM_white_paper_2.0.pdf> (last accessed 22 December 2008)

BRAUNING, D. W. [ED.]. 1992. Atlas of breeding birds in Pennsylvania. University of Pittsburgh Press, Pittsburgh, PA.

BUCKLAND, S. T., D. R. ANDERSON, K. P. BURNHAM, J. L. LAAKE, D. L. BORCHERS, AND L. THOMAS. 2004. Advanced distance sampling: estimating abundance of biological populations. Oxford University Press, UK.

CHERRILL, A., AND C. MCCLEAN. 1995. An investigation of uncertainly in field habitat mapping and the implications for detecting land cover change. Landscape Ecology 10:5–21.

CLEVENGER, A. P., J. WIERZCHOWSKI, B. CHRUSZCZ, AND K. GUNSON. 2002. GIS-generated, expert-based models for identifying habitat linkages and planning mitigation passages. Conservation Biology 16:503–514.

CONNOR, E. F., AND E. D. MCCOY. 1979. The statistics and biology of the species-area relationship. The American Naturalist 113:791–833.

CONROY, M. J., AND B. R. NOON. 1996. Mapping of species richness for conservation of biological diversity: conceptual and methodological issues. Ecological Applications 6:763–773.

CUTRIGHT, N. J., B. R. HARRIMAN, AND R. W. HOWE. 2006. Atlas of the breeding birds of Wisconsin. University of Wisconsin Press, Madison, WI.

Dean, D. J., K. R. Wilson, and C. H. Flather. 1997. Spatial error analysis of species richness for a GAP analysis map. Photogrammetric Engineering and Remote Sensing 63:1211–1217.

Delury, D. B. 1954. On the assumptions underlying estimates of mobile populations, pp. 287–293. *In* O. Kempthorne, D. R. Bancroft, J. W. Gowen, and J. L. Lush [eds.]. Statistics and mathematics in biology. Iowa State College, Ames, IA.

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

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

Goldstein, P. Z. 1999. Functional ecosystems and biodiversity buzzwords. Conservation Biology 13:247–255.

Goodchild, M., and S. Gopal [eds.]. 1989. The accuracy of spatial databases. Taylor and Francis, London. *Cited in* G. T. Koeln, L. M. Cowardin, and L. L. Strong. 1996. Geographic Information Systems, pp. 540–566. *In* T. A. Bookhout [ed.]. Research and management techniques for wildlife and habitats. The Wildlife Society, Bethesda, MD.

Great Basin Bird Observatory. 2005. Landbirds of Nevada and the habitats they need: A resource manager's guide to conservation priority species. Great Basin Bird Observatory Technical Report No. 05-01. Great Basin Bird Observatory, Reno, NV.

Green, K. 1994. The potential and limitations of remote sensing and GIS in providing ecological information, pp. 327–336. *In* V. A. Sample [ed.]. Remote sensing and GIS in ecosystem management. Island Press, Washington, DC.

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

Gurda, R. F. 1994. Linking and building institutions for a statewide landcover mapping program. GIS/LIS 1994 Proceedings 1:403–412.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773–785.

Hortal, J., J. M. Lobo, and A. Jiménez-Valverde. 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. Conservation Biology 21:853–863.

Johnsgard, P. A. 1973. Grouse and quails of North America. University of Nebraska Press, Lincoln, NE.

Kantrud, H. A., G. L. Krapu, and G. A. Swanson. 1989. Prairie basin wetlands of the Dakotas: a community profile. U.S. Fish and Wildlife Service, Washington, DC, USA. Biological Report 85(7.28).

Knopf, F. L. 1996. Prairie legacies—birds, pp. 135–148, *In* F. B. Samson and F. L. Knopf [eds.]. Prairie Conservation: preserving North America's most endangered ecosystem. Island Press, Washington, DC.

Krebs, C. J. 1989. Ecological methodology. Harper & Row, NY.

Lillesand, T. M., and R. W. Kiefer. 2000. Remote sensing and image interpretation. John Wiley & Sons, NY.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy Estimation and Modeling. Elsevier, Amsterdam.

McKelvey, K. S., K. B. Aubry, and M. K. Schwartz. 2008. Using anecdotal occurrence data for rare or elusive species: the illusion of reality and a call for evidentiary standards. BioScience 58:549–555.

Millspaugh, J. J., and F. R. Thompson, III [eds.]. 2008. Models for Planning Wildlife Conservation in Large Landscapes. Elsevier Science. Burlington, MA.

Newbold, S., and J. M. Eadie. 2004. Using species-habitat models to target conservation: a case study with breeding mallards. Ecological Applications 14:1384–1393.

Niemuth, N. D. 2003. Identifying landscapes for prairie chicken translocation using habitat models and GIS: a case study. Wildlife Society Bulletin 31:145–155.

Niemuth, N. D., M. E. Estey, and C. R. Loesch. 2005. Developing spatially explicit habitat models for grassland bird conservation planning in the Prairie Pothole Region of North Dakota, pp. 469–477. *In* C. J. Ralph and T. D. Rich [eds.]. Bird Conservation Implementation and Integration in the Americas: Proceedings of the Third International Partners in Flight Conference 2002 U.S. Department of Agriculture, Forest Service General Technical Report PSW-191.

Niemuth, N. D., R. E. Reynolds, D. A. Granfors, R. R. Johnson, B. Wangler, and M. E. Estey. 2008. Landscape-level Planning for Conservation of Wetland Birds in the U.S. Prairie Pothole Region, pp. 533–560. *In* J. J. Millspaugh and F. R. Thompson, III [eds.], Models for Planning Wildlife Conservation in Large Landscapes. Elsevier Science. Burlington, MA.

Noss, R. F. 1987. Do we really want diversity? Whole Earth Review 1987:126–128.

Rodda, G. H. 1993. How to lie with biodiversity. Conservation Biology 7:959-960.

Sauer, J. R., J. E. Hines, and J. Fallon. 2008. The North American Breeding Bird Survey, results and analysis 1966–2007. Version 5.15.2008. USGS Patuxent Wildlife Research Center, Laurel, MD. [Online.] <http://www.mbr-pwrc.usgs.gov/bbs/bbs.html> (22 December 2008).

Schroeder, M. A., and L. A. Robb. 1993. Greater prairie-chicken (*Tympanuchus cupido*), pp. 1–24. *In* A. Poole and F. Gill [eds.], The Birds of North America, No. 36. The Birds of North America, Inc., Philadelphia, PA.

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

Stein, B. A., C. Scott, and N. Benton. 2008. Federal lands and endangered species: the role of military and other federal lands in sustaining biodiversity. BioScience 58:339–247.

Thapa, K., and J. Bossler. 1992. Accuracy of spatial data used in geographic information systems. Photogrammetric Engineering and Remote Sensing 58:835–841.

Thogmartin, W. E., A. L. Gallant, M. G. Knutson, T. J. Fox, and M. J. Suárez. 2004. Commentary: a cautionary tale regarding use of the National Land Cover Dataset 1992. Wildlife Society Bulletin 32:970–978.

Thogmartin, W. E., M. G. Knutson, and J. R. Sauer. 2006a. Predicting regional abundance of rare grassland birds with a hierarchical spatial count model. Condor 108:25–46.

Thogmartin, W. E., F. P. Howe, F. C. James, D. H. Johnson, E. T. Reed, J. R. Sauer, and F. R. Thompson III. 2006b. A review of the population estimation approach of the North American Landbird Conservation Plan. Auk 123:892–904.

Villard, M. A., E. V. Schmidt, and B. A. Maurer. 1998. Contributions of spatial modeling to avian conservation, pp. 49–64. *In* J. M. Marzluff and R. Sallabanks [eds.], Avian conservation: research and management. Island Press, Washington, DC.

Westemeier, R. L., and S. Gough. 1999. National outlook and conservation needs for greater prairie chickens, pp. 169–187. *In* W. D. Svedarsky, R. H. Hier, and N. J. Silvy [eds.], The greater prairie chicken: a national look. University of Minnesota Agricultural Experiment Station Miscellaneous Publication 99-1999, Saint Paul, MN.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385–397.

Wiens, J. A. 2002. Predicting species occurrences: progress, problems, and prospects, pp. 739–749. *In* J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson [eds.], Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, DC.

Wilen, B. O., and M. K. Bates. 1995. The US Fish and Wildlife Service's National Wetlands Inventory project. Vegetatio 118:153–169.

Woodbury, P. B. 2003. Dos and don'ts of spatially explicit ecological risk assessments. Environmental Toxicology and Chemistry 22:977–982.