Conflation and aggregation of spatial data improve predictive models for species with limited habitats: A case of the threatened yellow-billed cuckoo in Arizona, USA

Miguel L. Villarreal a,*, Charles van Riper III b, Roy E. Petrakis c

a U.S. Geological Survey, Western Geographic Science Center, Menlo Park, CA 94025, USA
b U.S. Geological Survey, Southwest Biological Science Center, Sonoran Desert Research Station, Tucson, AZ 85721, USA
c University of Arizona, School of Geography and Development, Tucson, AZ 85719, USA

Keywords:
Avian habitat
Riparian
Yellow-billed cuckoo
Land-use/land-cover
NDVI
Geospatial data quality
Data aggregation
Species distribution model

Abstract

Riparian vegetation provides important wildlife habitat in the southwestern United States, but limited distributions and spatial complexity often leads to inaccurate representation in maps used to guide conservation. We test the use of data conflation and aggregation on multiple vegetation/land-cover maps to improve the accuracy of habitat models for the threatened western yellow-billed cuckoo (Coccyzus americanus occidentalis). We used species observations (n = 479) from a state-wide survey to develop habitat models from 1) three vegetation/land-cover maps produced at different geographic scales ranging from state to national, and 2) new aggregate maps defined by the spatial agreement of cover types, which were defined as high (agreement = all data sets), moderate (agreement > 2), and low (no agreement required). Model accuracies, predicted habitat locations, and total area of predicted habitat varied considerably, illustrating the effects of input data quality on habitat predictions and resulting potential impacts on conservation planning. Habitat models based on aggregated and conflated data were more accurate and had higher model sensitivity than original vegetation/land-cover, but this accuracy came at the cost of reduced geographic extent of predicted habitat. Using the highest performing models, we assessed cuckoo habitat preference and distribution in Arizona and found that major watersheds containing high-probably habitat are fragmented by a wide swath of low-probability habitat. Focus on riparian restoration in these areas could provide more breeding habitat for the threatened cuckoo, offset potential future habitat losses in adjacent watershed, and increase regional connectivity for other threatened vertebrates that also use riparian corridors.

Introduction

Riparian areas in southwestern North America occupy only a small fraction of total land area but provide habitat for exceptionally large number of terrestrial and aquatic species, several of which are considered threatened or endangered (Knopf, Johnson, Rich, Samson, & Szaro, 1988; Robinson, Tockner, & Ward, 2002; Tockner & Stanford, 2002). In these systems surface and subsurface water contribute to highly productive vegetation communities that are both temporally dynamic and spatially heterogeneous (Naiman & Decamps, 1997). Water diversion, dams, groundwater pumping for urban uses, and livestock grazing have put considerable pressure on riparian and riverine ecosystems, often causing severe degradation or complete collapse (Busch & Smith, 1995; Fleischner, 1994; Poff, Koestner, Neary, & Henderson, 2011), and climate-related temperature and water stresses are expected to exacerbate the ongoing effects of human land uses (Perry, Andersen, Reynolds, Nelson, & Shafroth, 2012). Riparian forests and woodlands provide habitat cohesion for all wildlife and especially migratory bird species traveling over fragmented landscapes. The loss of riparian habitat can have implications at both local and global scales, particularly for the conservation of Neotropical migrant birds (Faaborg et al., 2010; Knopf & Samson, 1994; Laymon, 1998; Skagen, Melcher, Howe, & Knopf, 1998).

The western yellow-billed cuckoo

The western yellow-billed cuckoo (Coccyzus americanus occidentalis) is a Neotropical migrant bird that forages and breeds almost exclusively in riparian areas (Laymon & Halterman, 1989).
The yellow-billed cuckoo winters in South America and begins its northbound migration around April when it travels to northern Mexico and the western United States to breed (Hughes, 1999; Sechrist, Paxton, Ahlers, Doster, & Rya, 2012). Yellow-billed cuckoos arrive in the Southwest in May and June, and their nesting activities peak between mid-July and early August (Corman & Wise-Gervais, 2005). Historically the cuckoo was distributed across a vast geographical area of the western United States, southern Canada, and northern Mexico (Hughes, 1999), but range retraction and population declines have been observed throughout the western breeding range over the past century, largely due to the loss and fragmentation of cottonwood-willow (Populus fremontii/Salix gooddingii) riparian habitat (Laymon & Halterman, 1989; Ohmart, 1994). The bird was once considered widespread and locally common in Arizona (Phillips, Marshall, & Monson, 1964; Swarth, 1914), and typically nested in mature riparian forests and woodlands in drainages along central and southern Arizona (Hamilton & Halterman, 1965).

Yellow-billed cuckoo declines as great as 96% have been noted in areas following major water impoundment projects (Laymon & Halterman, 1989; Ohmart, 1994). As a result of habitat loss, the western yellow-billed cuckoo is currently a candidate for listing under the Endangered Species Act (Federal Register, 2011, p. 66391, 2013, pp. 61621–61666). Specific threats to the bird that warrant its consideration under the Endangered Species Act include “habitat destruction, modification, and degradation from dam construction and operations; water diversions; riverflow management; stream channelization and stabilization; conversion to agricultural uses, such as crops and livestock grazing; urban and transportation infrastructure; and increased incidence of wildfire. These factors also contribute to fragmentation and promote conversion to non-native plant species, particularly tamarisk” (Federal Register, 2013, pp. 61621–61666). Arizona is believed to support the largest remaining cuckoo population of states west of the Rocky Mountains, and the bird is currently listed as “State Threated” (immense jeopardy of becoming endangered) in Arizona (Arizona Game and Fish Department, 2002, 5 pp.).

**Yellow-billed cuckoo habitat**

Information on yellow-billed cuckoo habitat requirements in the western United States is currently insufficient for accurate conservation assessment in the Southwest. Early research describing yellow-billed cuckoo habitat in northern California provide useful reference information (Gaines & Laymon, 1984; Laymon & Halterman, 1989), but geomorphological and ecological differences between riverine systems in the Southwest and northern California preclude the direct application of information from these studies for conservation in the arid Southwest. One similarity is cuckoo preference for cottonwood and willow riparian vegetation in both California and Arizona. Gaines (1974) and Gaines and Laymon’s (1984) studies in California indicated that habitat patch size was extremely important for yellow-billed cuckoo occupancy, and suggested the cuckoo preferred patches of cottonwood-willow greater than 100 m wide and greater than 10 ha in area. Laymon and Halterman (1989) updated that figure to riparian woodlands greater than 15 ha. However, large riparian patches may not be as important to yellow-billed cuckoo in Arizona, where most riparian corridors are constrained by topography and water and are of lower productivity and cover less area when compared to northern California (Johnson, Hatten, Holmes, & Shafroth, 2012, 53 pp.).

The extent to which yellow-billed cuckoo use upland vegetation, wetlands, and non-native riparian species is not well understood, but is critical to conserving and managing scarce riparian habitat in the Southwest. Recent studies have found the cuckoo and other Neotropical migrant birds may utilize non-native shrubs like saltcedar (Tamarix spp.) as marginal habitat when native species are not present (Sogge, Sferra, & Paxton, 2008), but other models suggest a negative relationship between saltcedar and cuckoo presence (Johnson et al., 2012, 53 pp.). Cuckoos also use non- riparian upland shrub/scrub habitat prior to breeding, and orchards and agriculture lands adjacent to streams may provide important food sources after their migration and prior to breeding (Hughes, 1999). The phenology of riparian and upland plants, particularly the timing and magnitude of vegetation green-up related to summer precipitation, appear to influence habitat selection by yellow-billed cuckoo (Wallace, Villarreal, & van Riper III, 2013). The selection of habitat based on phenology is likely related to production of food sources that include cicadas, katydids and caterpillars (Hughes, 1999).

**Yellow-billed cuckoo and riparian habitat models**

In addition to a shortage of information describing local and landscape habitat preference in Arizona, attempts to model and understand cuckoo habitat use at a regional scale are influenced by the type, scale and quality of input data, which can ultimately affect conservation prioritization (Roloff, Donovan, Linden, & Strong, 2009; Thogmartin, Gallant, Knutson, Fox, & Suárez, 2004; Wilson et al., 2005). Habitat estimation using only bioclimatic niche variables is insufficient for riparian-dependent species because the distribution of riparian vegetation is more influenced by water and land use (including water impoundment, groundwater pumping, effluent outflow, urbanization and the impacts of invasive species) than by climate and topographic variables (Glenn et al., 2001; Jones et al., 2010; Shafroth, Stromberg, & Patten, 2002; Smith, Devitt, Sala, Cleverly, & Busch, 1998; Villarreal, Drake, Marsh, & McCoy, 2012). Land-use/land-cover (LULC) and vegetation maps are, therefore, key environmental variables for modeling vertebrate habitat distribution at broad scales, especially in rare habitat types or habitat affected by adjacent land uses (Giordano, Navarro, & Martella, 2010; Jennings, 2000; Thullier, Araújo, & Lavoré, 2004).

The effect of input LULC data type, scale, and quality on output predictions have been carefully considered for a range of GIS modeling techniques (Atatian, Sederling, Heaton, & Blomberg, 2010; Flather, Wilson, Dean, & McComb, 1997; Fleming, Didier, Miranda, & Porter, 2004; Syphard, Clarke, Franklin, Regan, & Mcginnis, 2011). Model uncertainties due to data type and quality can become even more inflated when working with riparian-dependent species: because of its structural and spatial complexity riparian vegetation is notoriously difficult to map over large areas using moderate resolution earth observation satellite data (Akaşeh, Neale, & Jayanthi, 2008; Goetz, Wright, Smith, Zinecker, & Schaub, 2003; Villarreal, Van Leeuwen, & Rono-Leon, 2012; Yang, 2007). Furthermore, the use of small, isolated patches of riparian forest as stopover habitat by migrating birds in the Southwest has shifted the focus of riparian conservation towards a mixture of dense corridors and small patches over the landscape (Skagen et al., 1998). These small patches are difficult to detect using remote sensing methods and rare cover types like wetlands are, therefore, often mapped with low accuracy in regional and national land cover data sets (Gallant, 2009). The process of re-existing LULC and vegetation data sets to model distribution of riparian-dependent species should, therefore, be approached with great caution and awareness of data quality issues.

Studies describing data quality effects on habitat model results often shed light on model uncertainty by highlighting errors related to differences between input data sets, and the ensuing error (Glenn & Ripple, 2004; McDermid et al., 2009). Here we take a...
different approach and seek to exploit spatial and thematic commonalities of multiple existing LULC/vegetation data sets to create new cover information that may ultimately have more value for habitat modeling than any single source alone. This approach, referred to as conflation in the Geographic Information Sciences, was originally developed to consolidate, compare, and eliminate errors in large digital map files produced by the US Census Bureau and US Geological Survey (Lynch & Saalfeld, 1985; Saalfeld, 1988). Conflation generally refers to the merging and integration of features with varying accuracies to create a hybrid data set of higher accuracy than the source maps (Samal, Seth, & Cueto1, 2004; Sester, Anders, & Walter, 1998; Walter & Fritsch, 1999). Conflation processes have primarily been applied to multi-attribute vector maps, but there are recent studies of raster data conflation and satellite image/land cover map conflation, which has also been referred to as “harmonization” or “fusion” in the remote sensing literature (Jung, Henkel, Herold, & Churkina, 2006; Veregin, Sincák, & Kopco, 2000).

We believe that by conflating multiple, pre-existing LULC maps it is possible maximize confidence in the accuracy of an output habitat model and reduce over-fitting and over-mapping of potential habitat. The purpose of our study was to: 1) to develop a refined species distribution model of the yellow-billed cuckoo for the state of Arizona and identify vegetation, landscape, and patch

Fig. 1. Map of the study area, including location of yellow-billed cuckoo detections and study drainages.
variables that contribute to habitat selection, 2) to investigate the use of data conflation of multi-source and multi-scale geospatial LULC for habitat estimation over a large geographic area, and 3) use the best available model to examine the amount and spatial distribution of yellow-billed cuckoo habitat in Arizona. This study addresses the following questions ranging from broad to specific:

1. How much yellow-billed cuckoo habitat is in Arizona, and where is it distributed?
2. What are the primary vegetation and landscape components of yellow-billed cuckoo habitat?
3. Can we improve habitat model accuracy using multi-source, multi-scale input data?
4. How does data quality influence the modeled spatial distribution of habitat and what are the implications for conservation?

Material and methods

Study area

The study area encompassed the entire 295,260 km² state of Arizona, USA (Fig. 1). Arizona has three broad physiographic regions: 1) the Basin and Range system in the south, where runoff from numerous high elevation “sky island” mountain ranges drain into broad alluvial valleys, 2) the Central Highlands, a mountainous region with numerous small streams draining through steep valleys, and 3) the Colorado Plateau, a high desert landscape formed from uplifted, dissected, and eroded sedimentary rock. The state contains parts of 18 major watersheds and 11 major drainages including the Agua Fria, Bill Williams, Colorado, Gila, Little Colorado, Lower Gila, Salt, San Pedro, Santa Cruz, Verde, and Virgin Rivers. Elevation ranges from 23 m on the Colorado River near the southwestern corner of the state to 3852 m at Humphreys Peak in northern Arizona. Because yellow-billed cuckoos breed only in riparian areas we focused our modeling on major drainages in the state by buffering a streams layer by 4 km (Fig. 1).

Habitat modeling: environmental data

We developed eight habitat models, three with existing land-cover data sets, three with new data sets created by aggregating and conflating classes from these data, one using only the most accurate individual classes selected from any original land-cover or conflated data set, and the eighth with only a vegetation index calculated from satellite spectral data and topography data. In addition, we calculated at each pixel location the amount of cover (ha) in a 480 m radius (72 ha area) for each land-cover type; a focal area that has proven optimal from other yellow-billed cuckoo studies in Arizona (Johnson et al., 2012, 53 pp.). All eight models included four “constant” environmental variables that are described in the next section. All data layers were at native 30 m resolution except one (a vector map), which was resampled to 30 m resolution.

Constant environmental variables

Four environmental variables were used in all models: the Normalized Differenced Vegetation Index (NDVI), neighborhood NDVI within a 480 m radius (NDVI480), Digital Elevation Model (DEM), and a Topographic Position Index (TPI).

NDVI was calculated from Landsat 7 (ETM+) images acquired in 2000 and mosaicked by the Arizona Remote Sensing Center (ARSC) at the University of Arizona. NDVI was calculated from the multi-spectral data as:

\[ \text{Band 4 (near – infrared)} - \text{Band 3 (red)} \]

We calculated two NDVI metrics: 1) NDVI at each pixel location, and 2) the sum of NDVI values within a 480 m radius surrounding a pixel.

The DEM was a seamless elevation model of Arizona in meters that was acquired from the USGS National Elevation Dataset (NED). TPI was calculated from the elevation dataset as the difference between the cell elevation and the mean elevation within a 480 m radius around that cell. Negative TPI values indicate incised canyons, high TPI indicates mountain peaks.

Land-use/land-cover source data

Land-cover data used for habitat models were produced at three geographic scales ranging from National to State-level: National Land Cover Database (NLCD), Southwestern Regional Gap (SWReGAP) and Arizona Game and Fish Department (AZGFD) riparian map. These data were developed for different entities and purposes, but each contained a number of common riparian forest and wetland classes that we used in the modeling effort (Table 1).

Table 1

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>SWReGAP</th>
<th>NLCD</th>
<th>AZGFD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest</td>
<td>Broadleaf/Cottonwood Forest</td>
<td>Riparian Woodland and Shrubland; Lower Montane Riparian Woodland and Shrubland</td>
<td>Deciduous Forest</td>
<td>Cottonwood Willow; Mixed Broadleaf</td>
</tr>
<tr>
<td>Mesquite Invasive</td>
<td>Mesquite Bosque</td>
<td>Riparian Mesquite Bosque</td>
<td>NA</td>
<td>Mesquite</td>
</tr>
<tr>
<td>Wetland</td>
<td>Wetlands and Marsh</td>
<td>North American and West Emergent Marsh</td>
<td>Woody Wetlands; Emergent Herbaceous Wetlands</td>
<td>Wet Meadow; Marsh</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Agriculture</td>
<td>Agriculture</td>
<td>Pasture/Hay; Row Crops;</td>
<td>Agriculture</td>
</tr>
</tbody>
</table>

Regional: SWReGAP is a regional vegetation map of 5 southwestern states including Arizona, Colorado, Nevada, New Mexico and Utah (Lowry et al., 2007). The region was divided into 4–6
“mapping zones” and mapping responsibilities were distributed among a consortium of university and federal agencies. SWReGAP contains 125 land cover classes: 109 representing vegetation cover or “ecological systems” and 16 land-use types. The map was created by classifying ETM+ images acquired from 1999 to 2001 that were trained and validated with field data. Average overall accuracy for the five mapping zones in Arizona was 60.8%. Overall accuracy and class accuracy information are available for each mapping zone; however the accuracy of individual riparian vegetation classes in AZ mapping zones cannot be determined because of the low number of validation points used in those classes. SWReGAP land-cover data can be accessed at: http://earth.gis.usu.edu/swgap/landcover.html (last accessed 04/25/2013).

State: The AZGFD riparian vegetation map is vector vegetation map of 12 riparian and wetland classes mapped using a mixture of Landsat TM and Multiple Resolution Aerial Videography with an iterative method of field sampling followed by map updating. The map was developed between 1993 and 1994 by the AZGFD as part of the State’s “Waters-Riparian Protection Program” to identify riparian areas near perennial waters and intermittent streams. The final accuracy of the map is unknown, but early validation estimates by AZGFD conducted in certain watersheds range from 35 to 58% (Wentworthlund, 1997). The data are freely available from Arizona State Land Department, Arizona Land Resource Information System: http://www.azland.gov/alris/ (last accessed 04/25/2013). The AZGFD map polygons were converted to 30 m rasters for our study.

Data conflation and aggregation. NLCD, SWReGAP and AZGFD map classes were cross-walked into a common LULC classification scheme to facilitate data conflation and aggregation. The five-class land-cover scheme was based on assumptions of yellow-billed cuckoo habitat use measured during the 1998–1999 field surveys (Johnson et al., 2010). Classes were (1) Forest (broadleaf/cottonwood-willow forest), (2) Mesquite (mesquite bosque (forest)), (3) Invasive (invasive shrubland), (4) Wetland (wetlands and marsh), and (5) Agriculture (Table 1). Each new class was extracted from the source map as an independent presence/absence data layer.

We used GIS map algebra to aggregate and conflate the individual layers derived from the NLCD, SWReGAP and AZGFD data sets based on the level of class agreement at each pixel location: High-agreement (data conflation: mapped as class X by all 3 data sets), moderate-agreement (conflation/aggregation: mapped as class X by ≥2 data sets), and low-agreement (aggregation: mapped as class X by any data set). For example, when developing an aggregated Forest layer, if AZGFD and NLCD both mapped a pixel as “Forest” at a given location, and SWReGAP mapped it as “Invasive”, then the output pixel would be classified as “moderate-agreement forest,” and so forth (Fig. 2). In this case where there is no absolute agreement between the three layers, this same pixel location would be classified as “low-agreement invasive” in the aggregated “Invasive” data layer.

Independent validation of land-use/land-cover maps. We assessed the accuracy of each original vegetation/land cover data set and the new aggregated data layers. Accuracy assessment of mapped products from remotely sensed data is typically accomplished using an error matrix analysis that quantifies the proportion of mapped pixels that are correctly mapped relative to a “ground truth” dataset (Congalton, 1991). This was accomplished by comparing mapped class X with high-resolution aerial imagery for a series of randomly distributed validation points. Imagery used for validation included 2007 National Agriculture Imagery Program (NAIP; 1 m resolution) of Arizona and Google Earth imagery which offers both recent and past imagery (e.g. 1996, 2004, 2011, and 2012) for parts of Arizona. We assumed no major changes between the periods of the LULC maps and NAIP imagery acquisition; however when we encountered observed disagreements between mapped classes and NAIP imagery that may have been the result of LULC change we consulted historical imagery from Google Earth. For each random point, we assigned a cover class based on interpretation of the aerial imagery and later compared that with the mapped class at that location. We were confident in our ability to visually discriminate different forest classes using aerial imagery; recent mapping of southwestern riparian vegetation using aerial photo interpretation produced excellent results (e.g. kappa = 0.941, overall accuracy = 95%; Villarreal, Drake, et al., 2012). For each data set we created a confusion matrix to determine producer’s and user’s accuracy of each class, total accuracy and a kappa value. The number of points used to validate each riparian class ranged from 20 to 100 depending on the spatial extent of the class. Total validation points used per map ranged from 140 to 300.

Habitat modeling: species observation data

Yellow-billed cuckoo observations used to develop habitat models were collected in 1998 and 1999 for an Arizona state-wide survey (Johnson et al., 2010). These presence and absence data, over a decade old, have been used to identify some key breeding areas and important habitat requirement of the bird (Johnson et al., 2010), and more recently used to investigate the influence of land surface phenology on riparian habitat selection by the cuckoo (Wallace et al., 2013); but to date no state-wide habitat maps have been created from these data, leaving a gap in our knowledge of the probable distribution of this threatened bird and its habitat in Arizona.

Yellow-billed cuckoo survey sites (n = 107) were selected on federal and state lands based on expert opinion, museum samples and information from the scientific literature indicating probable habitat (Johnson et al., 2010). Survey sites targeted a total of 29 major river courses within the state, with more intensive surveys on the Bill Williams, Colorado, San Pedro, Santa Cruz, and Verde
River drainages (Fig. 1). Surveys were conducted from 05/31/1998 to 08/29/1998, and 06/09/1999 to 09/02/1999, between 0600 and 1300 h. Yellow-billed cuckoo are known to be a secretive and elusive species, therefore detections were based on taped calls that were broadcast (~ 100 m range) to illicit a response from the bird. Broadcasts were made every 100 m along the survey transects. If a cuckoo did not respond to the initial call, the broadcast was repeated a maximum of four times before moving 100 m to the next broadcast point location. The accuracy of any point location collected using the GPS receiver during the data collection period was approximately 30± meters and is based on averaging the total positional error for all points.

Surveys were conducted in 1998 (n = 97) and in 1999 (n = 196). We grouped detections from both years (n = 479) and randomly removed ~25% detections (n = 120) for independent model validation, leaving 359 detection points to train the MaxEnt model and estimate AUC. We grouped 1998 and 1999 absence locations (n = 244) (identified as transect start, mid-point, and stop locations) which were used to validate the model (but not to train it).

As an additional independent validation we use data from a recent (July–August, 2012) yellow-billed cuckoo repeat survey conducted on the Upper Santa Cruz River where we have comprehensive 2007 digitized vegetation distribution and past changes (Villarreal, Drake, et al., 2012). Our 2012 surveys were conducted using the same protocols as the 1998–1999 state-wide surveys. These surveys were conducted along a 22 km stretch of the Upper Santa Cruz River watershed in riparian habitats ranging from mixed shrub-herbaceous to cottonwood-willow gallery riparian forests.

Yellow-billed cuckoo habitat models

We used MaxEnt (version 3.3.3k), a robust and widely applied habitat model (Phillips, Anderson, & Schapire, 2006), to develop eight probabilistic species habitat models from the 1998 to 1999 yellow-billed cuckoo detection locations and the environmental layers described above. Three models used the six land cover layers from the unaltered original data plus the four constant environmental variables: NLCD model, SWReGAP model and AZGFD model. Three were from the six conflated and aggregated land cover layers, data plus the four constant environmental variables: 1) High Agreement (HA) model, 2) Moderated Agreement (MA), and 3) Low Agreement (LA). We also selected the most accurate individual classes from any of the original LULC data sets to conflated/aggregated maps to create one composite model (henceforth, “composite”), whose layers were combined with the four constant environmental variables. Finally we developed a constants-only model (henceforth, “constants model”) using only the four constant environmental variables. The high-agreement map, produced by conflating the three input data sets, was the most accurate riparian vegetation data set (overall accuracy = 75%; k = 0.69) while the less accurate was the low agreement map (47%) (Table 2). Individual class accuracies varied considerably; for example, forest class user’s accuracy ranged from 92% (high agreement) to 38% (SWReGAP and low-agreement) (Table 2). Because of the conservative nature of the conflation approach, the high-agreement map had a high user’s accuracy for forest (92%) but a relatively low producer’s accuracy (74%). This indicates the mapped forest class is accurate and tightly constrained to areas that are likely true forest, but that the high-agreement conflation approach results in errors of omission. Moderate-agreement forest class, on the other hand, had a more balanced user’s (80%) and producer’s (85%) accuracy, with a more equal distribution of commission and omission errors. NLCD had a high forest user’s accuracy (78%) and a lower producer’s (59%), which is likely related to the inclusion of invasive, mesquite and broadleaf forests into one forest class. Of the three original LULC data sets, AZGFD had highest overall accuracy and kappa. AZGFD forest class accuracy was relatively low (67% user’s and 85% producers), and the high overall and kappa values for AZGFD are related to the relatively high accuracy of invasive, mesquite and wetland classes. Accuracy of wetland classes was generally low for all data sets, and this contributed to lower overall accuracy and kappa scores of the source maps (Table 2).

We created the composite LULC map using the most accurate individual classes listed in Table 2: 1) AZGFD invasive, 2) AZGFD invasive480, 3) HA mesquite, 4) MA mesquite480, 5) HA wetland, 6) HA wetland480, 7) MA forest, 8) MA forest480, 9) SWReGAP.

Table 2 Overall accuracy, Kappa and Producer’s and User’s accuracy of vegetation data: 1) National Land Cover Dataset model — NLCD, 2) Southwest Regional GAP Analysis Project model, 3) Arizona Game and Fish Department model — AZGFD, 4) High Agreement model — HA, 5) Moderate Agreement model — MA, and 6) Low Agreement model — LA.

<table>
<thead>
<tr>
<th></th>
<th>NLCD</th>
<th>SWReGAP</th>
<th>AZGFD</th>
<th>HA</th>
<th>MA</th>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall accuracy</td>
<td>58%</td>
<td>49%</td>
<td>62%</td>
<td>75%</td>
<td>56%</td>
<td>47%</td>
</tr>
<tr>
<td>User’s accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>78%</td>
<td>38%</td>
<td>67%</td>
<td>92%</td>
<td>80%</td>
<td>38%</td>
</tr>
<tr>
<td>Mesquite</td>
<td>40%</td>
<td>50%</td>
<td>50%</td>
<td>61%</td>
<td>54%</td>
<td>54%</td>
</tr>
<tr>
<td>Wetland</td>
<td>10%</td>
<td>36%</td>
<td>50%</td>
<td>60%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Invasive</td>
<td>40%</td>
<td>40%</td>
<td>70%</td>
<td>64%</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>Agriculture</td>
<td>76%</td>
<td>85%</td>
<td>70%</td>
<td>89%</td>
<td>68%</td>
<td>68%</td>
</tr>
<tr>
<td>Producer’s accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>59%</td>
<td>74%</td>
<td>85%</td>
<td>74%</td>
<td>85%</td>
<td>83%</td>
</tr>
<tr>
<td>Mesquite</td>
<td>50%</td>
<td>63%</td>
<td>79%</td>
<td>53%</td>
<td>47%</td>
<td>47%</td>
</tr>
<tr>
<td>Wetland</td>
<td>80%</td>
<td>67%</td>
<td>70%</td>
<td>92%</td>
<td>67%</td>
<td>48%</td>
</tr>
<tr>
<td>Invasive</td>
<td>36%</td>
<td>68%</td>
<td>67%</td>
<td>54%</td>
<td>45%</td>
<td>45%</td>
</tr>
<tr>
<td>Agriculture</td>
<td>97%</td>
<td>83%</td>
<td>78%</td>
<td>97%</td>
<td>94%</td>
<td>89%</td>
</tr>
</tbody>
</table>
agriculture, 10) SWReGAP agriculture480, 11) DEM, 12) NDVI, 13) NDVI480, and 14) TPI.

Accuracy of MaxEnt habitat models

The top three MaxEnt models, which together accounted for the highest TSS, Kappa, AUC, and overall accuracy scores, were: Composite, NLCD and moderate agreement (MA) (Table 3). Model AUC values ranged from 96.3 (NLCD) to 92.1 (high agreement). Kappas ranged from 0.57 (composite) to 0.35 (high agreement), and overall accuracy ranged from 80% (NLCD and moderate agreement) to 73% (high agreement) (Table 3). The Composite model had the highest sensitivity (probability that the model will correctly classify habitat at a presence location) at 70% of presence locations correctly identified, but the MA model had the highest specificity (the probability that the model will correctly predict an absence) with 90% of the absence locations correctly identified. The MA model had the highest TSS, Kappa, AUC, and overall accuracy scores, were: Composite, NLCD and moderate agreement (MA) (Table 3). Model AUC values ranged from 96.3 (NLCD) to 92.1 (high agreement). Kappas ranged from 0.57 (composite) to 0.35 (high agreement), and overall accuracy ranged from 80% (NLCD and moderate agreement) to 73% (high agreement) (Table 3). The Composite model had the highest sensitivity (probability that the model will correctly classify habitat at a presence location) at 70% of presence locations correctly identified, but the MA model had the highest specificity (the probability that the model will correctly predict an absence) with 90% of the absence locations correctly identified as non-habitat (Table 3). Specificity (errors of commission) will be low if the model over-predicts habitat, and sensitivity (errors of omission) will be low if the model is under-predicting. An assessment of sensitivity and specificity for TSS values ranged from 0.56 (composite model) to 0.31 (high agreement) (Table 3). Interestingly, the constants model, which used no land cover data at all, had higher TSS, Kappa and AUC scores than the low agreement, AZGFD, SWReGAP and high agreement models.

MaxEnt variable contributions

Contributions of individual input variables to habitat prediction varied from model to model (Table 4). The high agreement conflated model had a largest number of variables contributing to the prediction (8), while the constants model depended heavily on two. The NLCD model is dominated by the Forest480 layer (a class that consisted of forest, mesquite and invasive (Table 1)), an input vegetation layer which received a very low producer’s accuracy during independent map validation. The dominant contributing variable to the constants model was NDVI480, which likely resulted in high estimates of potential habitat (Table 4). The composite model relied heavily on NDVI480 and DEM, but also had contributions from DEM, Forest480 and Mesquite480, which likely limited the over-mapping of habitat for this model when compared to the constants model. The moderate agreement model had a balanced contribution of vegetation and constant variables (Table 4).

Yellow-billed cuckoo habitat preference

Comparison of average variable values at bird presence and absence locations for the three best models (composite, MA and NLCD) indicates that locations where birds were present were slightly higher elevations (1042 vs. 990 m), in moderately incised canyons, had more green vegetation (higher NDVI), and had more hectares of forest and mesquite than locations with no birds (Table 5). Cuckoos were found in areas with less wetland when compared to the absence points (Table 5). The MA and NLCD models show less wetland in areas where cuckoos are present vs. absent. Wetlands do not appear to be a factor in presence or absence locations for the composite model, but this is likely related to under-mapping of these types (Composite model used the HA input data layer). The Composite and Moderate Agreement models indicate invasive species cover was slightly greater at locations where cuckoos were detected (Table 5). The composite model, using the SWReGAP Agriculture class suggests that areas with cuckoos have less agriculture than areas with no cuckoos, but the MA and NLCD showed the opposite with more agriculture at cuckoo presence locations (Table 5). However, because invasive, agricultural, and wetland variables were not major contributors to the three habitat models, these differences may be insignificant habitat factors for cuckoos.

Amount and distribution of yellow-billed cuckoo habitat in Arizona

State-wide estimates of high probability habitat ranged from 24,290 ha (AZGFD) to 64,323 ha (constants model). The three most accurate models predicted 33,929 ha (composite), 29,867 ha (moderate agreement), and 36,392 (NLCD) for the

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables (n)</th>
<th>Percent contribution (ranked)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>14</td>
<td>NDVI480 (24.6), Forest480 (23.1), DEM (15.7), Forest (15.4), Mesq480 (6.3)</td>
</tr>
<tr>
<td>MA</td>
<td>14</td>
<td>NDVI480 (23.5), Forest480 (16.7), DEM (15.7), Mesq480 (9.9), NDVI (8.4), TPI (8.2)</td>
</tr>
<tr>
<td>HA</td>
<td>14</td>
<td>Forest480 (22.3), DEM (17.2), NDVI480 (13.7), TPI (11.5), Agri480 (8.8), NDVI (7.4), Mesq480 (6.20), Forest (6.1)</td>
</tr>
<tr>
<td>Composite</td>
<td>14</td>
<td>NDVI480 (52.1), DEM (22.7), Forest480 (8.3), Mesq480 (6.6)</td>
</tr>
<tr>
<td>AZGFD</td>
<td>10</td>
<td>Mesq480 (66.6), AZGFD_LC (9.8), Agri480 (6)</td>
</tr>
<tr>
<td>SWReGAP</td>
<td>10</td>
<td>SWReGAP_LC (24.2), Agri480 (15.7), NDVI480 (14.5), Forest480 (14), DEM (11.9), NDVI (6.5), Mesq480 (6.1)</td>
</tr>
<tr>
<td>NLCD</td>
<td>8</td>
<td>Forest480 (63.2), DEM (10.8), TPI (7.2), NLCD_LC (5.4), NDVI (5.2)</td>
</tr>
<tr>
<td>Constants</td>
<td>4</td>
<td>NDVI480 (74.9), DEM (12.2), TPI (8)</td>
</tr>
</tbody>
</table>

Table 4
Number of input variables and percent contribution for 1) Low Agreement model — LA, 2) Moderate Agreement model — MA, 3) High Agreement model — HA, 4) Composite model, 5) Arizona Game and Fish Department model, 6) Southwest Regional GAP Analysis Project model, 7) National Land Cover Dataset model, and 8) High Agreement — HA model.

---

Table 3
Matrix of habitat model accuracy results. True Skill Statistic (TSS), Kappa, Area Under the Curve (AUC), sensitivity, specificity, Producer’s accuracy and 2012 survey accuracy ranked by TSS, for 1) Composite model, 2) National Land Cover Dataset model, 3) Moderate Agreement model — MA, 4) Constants model, 5) Low Agreement model — LA, 6) Arizona Game and Fish Department model — AZGFD, 7) Southwest Regional GAP Analysis Project model, and 8) High Agreement — HA model.

<table>
<thead>
<tr>
<th>Model</th>
<th>TSS</th>
<th>Kappa</th>
<th>AUC</th>
<th>Overall accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Producer’s accuracy</th>
<th>Producer’s non-habitat</th>
<th>2012 survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>0.56</td>
<td>0.57</td>
<td>95.1</td>
<td>81</td>
<td>70</td>
<td>86</td>
<td>72</td>
<td>85</td>
<td>88</td>
</tr>
<tr>
<td>NLCD</td>
<td>0.52</td>
<td>0.54</td>
<td>96.3</td>
<td>80</td>
<td>63</td>
<td>87</td>
<td>72</td>
<td>84</td>
<td>63</td>
</tr>
<tr>
<td>MA</td>
<td>0.49</td>
<td>0.52</td>
<td>95.4</td>
<td>80</td>
<td>59</td>
<td>90</td>
<td>74</td>
<td>82</td>
<td>63</td>
</tr>
<tr>
<td>Constants</td>
<td>0.47</td>
<td>0.47</td>
<td>95.3</td>
<td>77</td>
<td>63</td>
<td>83</td>
<td>65</td>
<td>82</td>
<td>63</td>
</tr>
<tr>
<td>LA</td>
<td>0.43</td>
<td>0.44</td>
<td>94.9</td>
<td>76</td>
<td>59</td>
<td>84</td>
<td>65</td>
<td>81</td>
<td>25</td>
</tr>
<tr>
<td>AZGFD</td>
<td>0.42</td>
<td>0.44</td>
<td>94.6</td>
<td>76</td>
<td>57</td>
<td>86</td>
<td>66</td>
<td>80</td>
<td>25</td>
</tr>
<tr>
<td>SWReGAP</td>
<td>0.41</td>
<td>0.44</td>
<td>94.2</td>
<td>77</td>
<td>53</td>
<td>89</td>
<td>60</td>
<td>79</td>
<td>25</td>
</tr>
<tr>
<td>HA</td>
<td>0.32</td>
<td>0.35</td>
<td>92.1</td>
<td>73</td>
<td>45</td>
<td>87</td>
<td>64</td>
<td>76</td>
<td>63</td>
</tr>
</tbody>
</table>

*Highest performing models.*
All models generally predicted the most high probability habitat in the Upper and Lower Verde and Upper and Lower San Pedro watersheds (Figs. 3 and 4; Supplemental figure online for maps of all models). It was in these two watersheds that Johnson et al. (2012, 53 pp.) found the greatest number of yellow-billed cuckoos. The amount of high probability cuckoo habitat within watersheds varied considerably from model to model particularly in watersheds that appeared to have the most high probability habitat (Fig. 4). The constants-only model predicted more hectares of habitat than the average of all models in every watershed except the Big Sandy (Fig. 4). NLDC and SWReGAP models, based on original source data, both had a number of watersheds where the predictions were much higher than the average; conversely AZGFD predictions were below the average for all but two watersheds (Fig. 5). When all models are averaged, convergence of high probability habitat occurs within a contiguous band of watersheds stretching west from southeastern to west-central Arizona (Fig. 3). The Composite Model indicates a distinct spatial fragmentation of habitat-rich watersheds when compared with the average of all models (Fig. 3).

### Table 5

<table>
<thead>
<tr>
<th>Variable</th>
<th>Composite</th>
<th>Moderate agreement</th>
<th>NLCD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Presence</td>
<td>Absence</td>
<td>Presence</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>Agriculture (ha)</td>
<td>2.52</td>
<td>8.01</td>
<td>6.98</td>
</tr>
<tr>
<td>DEM (m)</td>
<td>1036</td>
<td>289</td>
<td>1036</td>
</tr>
<tr>
<td>Forest480 (ha)</td>
<td>3.43</td>
<td>4.73</td>
<td>3.18</td>
</tr>
<tr>
<td>Invasive (ha)</td>
<td>1.78</td>
<td>8.54</td>
<td>2.33</td>
</tr>
<tr>
<td>Mesquite (ha)</td>
<td>0.71</td>
<td>2.40</td>
<td>0.55</td>
</tr>
<tr>
<td>NDVI</td>
<td>-0.10</td>
<td>0.14</td>
<td>-0.10</td>
</tr>
<tr>
<td>NDVI480</td>
<td>0.50</td>
<td>0.21</td>
<td>0.50</td>
</tr>
<tr>
<td>Wetland (ha)</td>
<td>0.00</td>
<td>0.03</td>
<td>2.17</td>
</tr>
</tbody>
</table>

### Discussion

Data aggregation and conflation increase model accuracy and applicability

Accurate, project-specific LULC and vegetation data require considerable field work and a high level of expertise to create, and can be prohibitively expensive when produced for large areas. Conservation and restoration decisions are, therefore, often based on models made from freely-available but often incomplete data that may not have been created at optimal geographic, thematic and spatial resolutions. The results of our research illustrated how habitat estimates made from existing LULC data sets varied considerably in both the amount of habitat predicted and the location. Data conflation and aggregation techniques require very little GIS expertise and can be easily achieved in study areas with multiple spatial data sources of varying accuracy, offering managers a cost-effective and more complete picture of possible habitat distributions. Our conflation and aggregation of three pre-existing data sets produced more accurate, high-resolution LULC maps.
Habitat models made from the aggregated and conflated maps, on the other hand, were informed by multiple input variables, allowed for more complex spatial relationships between input variables (e.g. a single detection location might occur in an area mapped as both high agreement forest and moderate agreement mesquite), and offered more information on factors contributing to yellow-billed cuckoo habitat selection. In addition to increased model complexity, one practical advantage of using habitat models created from aggregated data for conservation planning is the ability to develop multiple estimates of species habitat distributions that have a range of model specificity and sensitivity: depending on what level of certainty is required for a given conservation decision, managers can use habitat models developed from different levels of aggregated data (with different accuracy) to answer certain questions or to accomplish different management goals. For example, a habitat model made from the moderate-agreement input map, which had low model sensitivity but high model specificity, might be an appropriate layer to identify a wide range of potential sites for future yellow-billed cuckoo habitat restoration or monitoring projects given the potential for species distribution shifts to marginal habitat use under future climate changes. Models made from direct conflation of multiple data sets, like the high-agreement model, provide a higher level of certainty and specificity and might be used in more “high-stakes” situations. For example, if the yellow-billed cuckoo is ultimately listed under the ESA, the high-agreement model can provide a highly accurate baseline of habitat locations for initial conservation and monitoring efforts.

### Vegetation information

Vegetation information collected during the 1989–99 field surveys indicate that yellow-billed cuckoo are predominantly found at locations with a cottonwood-willow overstory and mixed understory dominated by willow (Johnson et al., 2010). Analysis of focal statistics indicate that the average amount of cottonwood-willow within 480 m of the detection points was more than twice that at absence locations; however, this amount (3.18–5.71 ha) was still much lower than estimates from studies in California (i.e. 10–15 ha; Gaines & Laymon’s, 1984; Laymon & Halterman, 1989). These focal estimates are clearly influenced by the accuracy and spatial distribution of the vegetation input data; focal estimates of vegetation cover measured from low-agreement aggregated maps provides a more liberal assessment of cover than conflated, high-agreement data. For example, the composite model, which contained the high-agreement mesquite layer, indicated less than 1 ha (average) of mesquite surrounding a cuckoo detection while the moderate agreement layer predicted 9.55 ha; the respective accuracy of the mesquite input layers for those models were 79% and 53%.

Comparison of yellow-billed cuckoo presence and absence locations indicated that sites with birds had more invasive vegetation in the surrounding 72 ha area than did sites with no birds (Table 5). This is a somewhat unexpected finding given that invasive plants like *Tamarix* are often implicated in the degradation of riparian systems and associated wildlife habitat (Shafroth et al., 2005). Recent studies indicate that a number of birds have adapted to, and thrived in, the invaded riparian communities of the Southwest (Hunter, Ohmart, & Anderson, 1988; Sogge et al., 2008; Van Riper, Paxton, O’Brien, Shafroth, & McGrath, 2008), and in some areas total volume and structure of riparian plant communities can be a greater influence on bird habitat use despite the dominance of non-native vegetation (Fleishman et al., 2003, Paxton, Van Riper, & O’Brien, 2008). This may be the case with the yellow-billed cuckoo where patches of invasive vegetation, when interspersed
with patches of native cottonwood-willow and mesquite forests, have greater foliage volume which may increase the total volume of available food sources.

NDVI has proven to be an adequate variable alone for predicting species habitat (Lahoz-Monfort, Guillera-Arroita, Milner-Gulland, Young, & Nicholson, 2010). The importance of NDVI as a variable in the yellow-billed cuckoo models, especially in the constants model, may be related to the high contrast between the NDVI values of productive floodplain vegetation and surrounding non-habitat upland desert, which limits its predictive power and contributes to low model sensitivity. However, NDVI480 was an important variable for MaxEnt models made from conflated/aggregated data, which suggests that NDVI also helps to discriminate between conflated classes, particularly Forest and Mesquite which were both contributing variables to the moderate agreement and composite model (Table 4). In addition to within-habitat discrimination, comparison of presence and absence locations, both of which were surveyed within comparable riparian areas indicates the cuckoos were found in sites that were slightly “greener” and had more surrounding green pixels than sites with no cuckoo observations (Table 5), a finding consistent with recent research documenting yellow-billed cuckoo selection of forest patches by greenness and phenology (Wallace et al., 2013).

Cuckoos were also found in areas with less entrenched floodplains when compared to absence locations (Table 5). Steep canyons typically have less area available for floodplain vegetation to grow and lack subtle floodplain terraces, topographic features which are associated with dense mesquite bosque habitat in open

---

**Fig. 5.** Deviation from average high probability yellow-billed cuckoo habitat within watersheds: A) National Land Cover Dataset model, B) Southwest Regional GAP Analysis Project model, C) Arizona Game and Fish Department model, D) Low Agreement model, E) Moderate Agreement model, F) High Agreement model, G) Composite model, and H) Constants model.
floodplains. The cuckoo’s presence in less entrenched streams also suggests that microhabitat may play a role in cuckoo habitat selection, where deeper and more incised canyons are typically more shaded and cooler than open floodplains, perhaps limiting forage productivity.

Agriculture did not contribute significantly to any of the models, but interestingly, comparison of agriculture cover around presence and absence points revealed that for some models yellow-billed cuckoo tended to occupy areas with more agriculture (Table 5). This could be the result of autocorrelation; sites with productive riparian vegetation would also be productive for crops, but more likely the yellow-billed cuckoo forage on insects from sources like crop lands, pastures or herbaceous vegetation surrounding the riparian corridor. While cuckoos primarily forage within cottonwood habitat, the use of adjacent lands for foraging may be related to the productivity of food sources (large insects) in these ecosystems and in response to precipitation from the southwestern monsoon (Wallace et al., 2013).

Conservation implications for sustaining cuckoo populations

To date there has been no spatial assessment of yellow-billed cuckoo habitat in the Southwest at any scale greater than individual watershed or river. Yellow-billed cuckoo generally have low site fidelity and can move among habitats seasonally or breed in different locations from year to year (Hughes, 1999), so a comprehensive and accurate habitat inventory is key for conservation planning. Large-scale habitat distribution information is important for conservation of migratory birds in general, as it provides information to prioritize conservation and restoration efforts in a spatially balanced way that helps reduce the risk of local extinction if key riparian areas are lost to land use changes, catastrophic flooding, or drought.

Nearly all of the models indicate an east-west band of low-probability habitat (primarily the Lower Gila, Middle Gila and Upper and Lower Salt River watersheds) that divides two core areas of high-probability habitat (the San Pedro and Verde watersheds), effectively fragmenting the north from the south (Fig. 3; See Supplemental figure online for maps of all models). This fragmentation is in part related to land-cover change and groundwater use in and around the Phoenix metropolitan area that extirpated riparian habitat over the last several decades. Prioritization of habitat restoration along river reaches within these low-probability watersheds could strengthen cuckoo populations in the state of Arizona and ameliorate losses if watersheds with abundant habitat faced declining conditions. For example, the San Pedro River was identified as containing a substantial amount of high yellow-billed cuckoo habitat, but the riparian habitat is in jeopardy. The San Pedro is one of the last free-flowing rivers in the Southwest whose riparian area now faces potentially catastrophic losses due to a recent Arizona Department of Water Resources decision to allow increased groundwater pumping to support urban growth in nearby Sierra Vista (Karp, 2013). If these changes were to materialize, small riparian restoration and conservation efforts in adjacent and more rural watersheds identified as important with our models might possibly offset these potential losses to yellow-billed cuckoo habitat on the San Pedro.

Acknowledgements

This research was made possible by a U.S. Geological Survey Mendenhall Research Fellowship supported by the Land Change Science and Land Remote Sensing Programs. Financial assistance was also provided by the U.S. Geological Survey Ecosystem and Environmental Health Mission Areas and the U.S. Fish and Wildlife Service via the Landscape Conservation Cooperatives program. In particular, we thank Susan Benjamin, Pattie Bright, David Lytle, Timothy Newman and Jonathan Smith for their continued support of our research. We also thank Stuart Marsh and Kyle Hartfield of the Arizona Remote Sensing Center (University of Arizona) for providing the Landsat mosaic of Arizona, and Ellis Margolis (University of Arizona, Laboratory of Tree Ring Research) for discussions contributing to these data improvement approaches. Terry Sohl and two anonymous reviewers provided constructive and insightful comments on early drafts.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apgeog.2013.12.003.

References

Arizona Game and Fish Department (2002). Western yellow-billed cuckoo (Coccyzus americanus occidentalis). Unpublished abstract compiled and edited by the Heritage Data Management System, Arizona Game and Fish Department, Phoenix, AZ.
Giordano, P. F., Navarro, J. L., & Martella, M. B. (2010). Building large-scale spatially explicit models to predict the distribution of suitable habitat patches for the


