

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

USDA National Wildlife Research Center - Staff  
Publications

U.S. Department of Agriculture: Animal and  
Plant Health Inspection Service

---

2018

## Accounting for observation processes across multiple levels of uncertainty improves inference of species distributions and guides adaptive sampling of environmental DNA

Amy J. Davis

*USDA National Wildlife Research Center, Amy.J.Davis@usda.gov*

Kelly E. Williams

*University of Washington*

Nathan P. Snow

*USDA APHIS Wildlife Services*

Kim M. Pepin

*USDA APHIS Wildlife Services*

Antoinette J. Piaggio

*USDA APHIS Wildlife Services*

Follow this and additional works at: [https://digitalcommons.unl.edu/icwdm\\_usdanwrc](https://digitalcommons.unl.edu/icwdm_usdanwrc)



Part of the [Life Sciences Commons](#)

---

Davis, Amy J.; Williams, Kelly E.; Snow, Nathan P.; Pepin, Kim M.; and Piaggio, Antoinette J., "Accounting for observation processes across multiple levels of uncertainty improves inference of species distributions and guides adaptive sampling of environmental DNA" (2018). *USDA National Wildlife Research Center - Staff Publications*. 2224.

[https://digitalcommons.unl.edu/icwdm\\_usdanwrc/2224](https://digitalcommons.unl.edu/icwdm_usdanwrc/2224)

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Animal and Plant Health Inspection Service at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in USDA National Wildlife Research Center - Staff Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

## ORIGINAL RESEARCH

# Accounting for observation processes across multiple levels of uncertainty improves inference of species distributions and guides adaptive sampling of environmental DNA

Amy J. Davis<sup>1</sup>  | Kelly E. Williams<sup>2</sup> | Nathan P. Snow<sup>1</sup> | Kim M. Pepin<sup>1</sup> | Antoinette J. Piaggio<sup>1</sup> 

<sup>1</sup>United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Fort Collins, Colorado

<sup>2</sup>School of Environmental and Forest Sciences, University of Washington, Seattle, Washington

## Correspondence

Amy J. Davis, United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Fort Collins, CO.  
Email: amy.j.davis@aphis.usda.gov

## Abstract

Understanding factors that influence observation processes is critical for accurate assessment of underlying ecological processes. When indirect methods of detection, such as environmental DNA, are used to determine species presence, additional levels of uncertainty from observation processes need to be accounted for. We conducted a field trial to evaluate observation processes of a terrestrial invasive species (wild pigs- *Sus scrofa*) from DNA in water bodies. We used a multi-scale occupancy analysis to estimate different levels of observation processes (detection,  $p$ ): the probability DNA is available per sample ( $\theta$ ), the probability of capturing DNA per extraction ( $\gamma$ ), and the probability of amplification per qPCR run ( $\delta$ ). We selected four sites for each of three water body types and collected 10 samples per water body during two months (September and October 2016) in central Texas. Our methodology can be used to guide sampling adaptively to minimize costs while improving inference of species distributions. Using a removal sampling approach was more efficient than pooling samples and was unbiased. Availability of DNA varied by month, was considerably higher when water pH was near neutral, and was higher in ephemeral streams relative to wildlife guzzlers and ponds. To achieve a cumulative detection probability >90% (including availability, capture, and amplification), future studies should collect 20 water samples per site, conduct at least two extractions per sample, and conduct five qPCR replicates per extraction. Accounting for multiple levels of uncertainty of observation processes improved estimation of the ecological processes and provided guidance for future sampling designs.

## KEYWORDS

detection, environmental DNA, invasive species, monitoring, multi-scale occupancy, *Sus scrofa*

## 1 | INTRODUCTION

A primary challenge to understanding ecological processes and the patterns they produce (e.g., survival, abundance, distribution) is that

they are rarely, if ever, observed perfectly. Understanding variation in the ability to detect a target species in the wild is necessary for disentangling the noise in observation processes (i.e., detection) from the signal of ecological processes of interest. The need to account

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published 2018. This article is a U.S. Government work and is in the public domain in the USA. *Ecology and Evolution* published by John Wiley & Sons Ltd.

for observation processes is well recognized in ecology, and many factors are known to impact these processes including observer error (Nichols, Hines, Sauer, Fallon, & Heglund, 2000), environmental conditions (Simons, Alldredge, Pollock, & Wettroth, 2007), detection method (Digby, Towsey, Bell, & Teal, 2013), and species behavior (Diefenbach, Marshall, Mattice, & Brauning, 2007). In molecular ecology, there are additional levels of uncertainty in observation processes dealing not just with species availability and human observation error but also with DNA availability and assay specificity and sensitivity (Willoughby, Wijayawardena, Sundaram, Swihart, & DeWoody, 2016). To effectively use molecular techniques to make inference to underlying ecological processes, it is necessary to evaluate and account for the various levels of uncertainty in observation processes (Bohmann et al., 2014; Hunter et al., 2015; McClintock et al., 2010; Spear, Groves, Williams, & Waits, 2015). Because the sources of error for genetic methods span multiple levels of biological organization, the overall error structure should be hierarchical—that is total error should be equivalent to a series of conditional probabilities.

Detection rates of individuals in the field may be influenced by vegetation or weather conditions such as dense foliage or rain that create visual obstructions, or road or stream noise that cause auditory disturbance, or time of day or season that may cause behavioral differences in detection rates (Christy, Yackel Adams, Rodda, Savidge, & Tyrrell, 2010; Farnsworth et al., 2002). Observation processes for molecular tools are unique because there are additional levels of observation (e.g., availability of DNA, DNA capture rate, and amplification success). Environmental conditions (pH, temperature, substrate, etc.) may strongly influence the observation processes at these different levels. Incorrectly accounting for multi-level influences could lead to biases in estimates of the underlying ecological process (Gu & Swihart, 2004). It is possible to get unbiased estimates of the cumulative detection process without accounting for the different levels of uncertainty (Schmidt, Kéry, Ursenbacher, Hyman, & Collins, 2013). However, identifying the observational processes by level allows for the optimization of sampling effort to increase overall detection probability. Thus, an understanding of both the factors affecting observation as well as the level they act on is critical for accurately quantifying ecological processes such as species distribution (invasive or endangered) or pathogen spread using molecular methods such as environmental DNA (eDNA; DNA collected from the environment rather than directly from a target species).

Detection of a target species' eDNA from water bodies is emerging as a potentially valuable method to infer the distribution of species and pathogens (Bohmann et al., 2014; Hunter et al., 2015; Takahara, Minamoto, & Doi, 2013). Efforts have been largely focused on aquatic and semi-aquatic species such as amphibians (Biggs et al., 2015; Pilliod, Goldberg, Arkle, & Waits, 2013; Schmidt et al., 2013), reptiles (Hunter et al., 2015; Piaggio et al., 2014), invertebrates (Doi et al., 2017; Thomsen et al., 2012), and fish (Takahara et al., 2013; Thomsen et al., 2012). As more recent work has applied eDNA detections in water bodies to terrestrial mammals (Rodgers & Mock, 2015; Ushio et al., 2017; Williams, Huyvaert, & Piaggio, 2017),

the method could be especially useful for detecting new invasions of terrestrial species in the early stages. To maximize the utility of eDNA in understanding landscape-level ecological processes (e.g., occupancy, distribution), it is important to determine the observation processes (i.e., detection rates) at many levels (Willoughby et al., 2016) that may influence the probability of detecting DNA in the environment. Once cells are shed, abiotic and biotic factors begin to degrade DNA (Barnes & Turner, 2016). Previous work has shown that the influence of microbial communities, temperature, pH, UV, and other environmental factors will impact the availability of DNA in the environment (Barnes & Turner, 2016) and therefore should be accounted for when assessing the detection probability of eDNA.

Occupancy models (MacKenzie et al., 2006) are well suited for quantifying species distributions in space and time while accounting for levels of uncertainty across observation processes, and have been used to assess species presence through eDNA (Hunter et al., 2015; Schmelzle & Kinziger, 2016; Schmidt et al., 2013; Valentini et al., 2016). When using eDNA as a passive detection method, there is the added complexity over a classical occupancy model because, even when the species of interest is present, the DNA in a given sample may not be present (Furlan, Gleeson, Hardy, & Duncan, 2016; Williams et al., 2017). Therefore, the detection process ( $p$ ) of a species by eDNA can be split into three levels: the probability that DNA is present and can be detected, "available", in the water sample ( $\theta$ ), the probability of capturing DNA in an extraction procedure ( $\gamma$ ) given it is available in the sample, and the probability of amplifying DNA in a qPCR run ( $\delta$ ) given it has been captured in an extract. Separating these probabilities allows evaluation of factors that influence observation of DNA across each of these levels as well as their influence on overall detection probability. By separating the observation process out in this manner, we can identify the level that will improve the most by increased sampling and thus ensure resources are optimally allocated. Enhancing overall detection probability will result in better inference of underlying ecological processes using this approach for the detection of cryptic (Bickford et al., 2007) or elusive (Rogala et al., 2011) species.

We sampled for eDNA of wild pigs (*Sus scrofa*), an invasive terrestrial species in North America and other parts of the world. They are capable of rapid geographic expansion (Snow, Jarzyna, & VerCauteren, 2017) and cause high levels of damage to ecosystems and the economy (Anderson, Sloomaker, Harper, Holderieath, & Shwiff, 2016; Chavarria, Lopez, Bowser, & Silvy, 2007; West, Cooper, & Armstrong, 2009). Because they are reliant on water bodies for drinking and wallowing, and their distribution and densities vary widely, they are a good model system for evaluating the potential application of eDNA for assessing the presence of terrestrial species, understanding spatial expansion of invasive species, and developing protocols for monitoring the effectiveness of invasive species control programs. As elimination programs for wild pigs occur in many countries across the globe, there is great need for cost-effective methods for evaluating success and guiding decisions (Hone, 1983; Korn & Bomford, 1996; Saunders & Bryant, 1988). Further, in areas without wild pigs, the ability to evaluate reports of sightings or sign is critical to implementation of early control measures that could curb the establishment of a newly

invasive population. Environmental DNA is a promising tool to aid in these monitoring efforts because it has the potential to be an efficient field method (Williams et al., 2017). Yet, application of eDNA in terrestrial species is currently limited by a poor understanding of the observation processes and inferences of ecological processes.

Our objectives were to: (a) examine factors that influence observation processes across several levels: availability of eDNA in water sampling, the capture rate of DNA in the extraction process, and the amplification probability during qPCR, (b) evaluate our ability to correctly assess target species presence at sampling sites given our observation process, and (c) develop an adaptive approach to eDNA collection and analysis to balance field and laboratory effort for efficiency. By accounting for multiple levels of uncertainty in the observation process, we aim to improve estimation of ecological processes and provide guidance for future sampling designs using eDNA for detection of a target species.

## 2 | METHODS

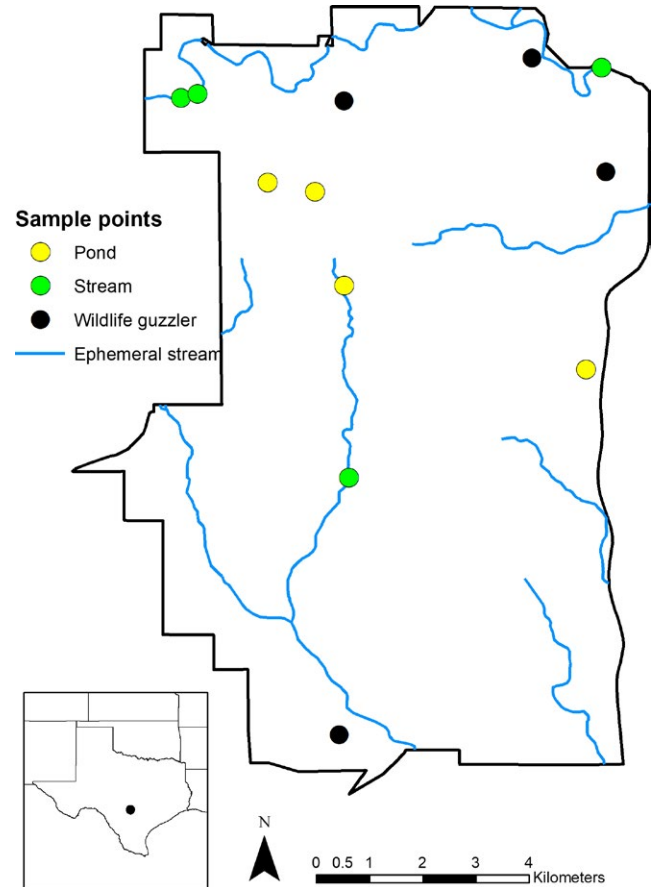
### 2.1 | Study area

One way to understand the detection probability of a method, such as eDNA, is to evaluate the ability to detect a target species in a setting where the species is known to be present. Typically, the presence or absence of a species is of primary interest and the detection probability is often thought of as a nuisance parameter that must be accounted for to obtain unbiased estimates of species occupancy (MacKenzie et al., 2002). However, to get precise estimates of detection probability based on the detection method alone, we can reframe the problem to eliminate the “nuisance” parameter of occupancy probability (MacKenzie et al., 2006) by sampling in an area where the presence of the species is known. This allows for the assessment of factors that might influence detection such as environmental (abiotic and biotic) factors or laboratory processes.

Our study was conducted at Camp Bullis Training Site (112.9 km<sup>2</sup>), in northern Bexar County, Texas, USA operated by Joint Base San Antonio (Figure 1). Camp Bullis is a restricted access property with perimeter fencing and high densities of pigs. This property is located in the Edwards Plateau ecoregion of the south-central semi-arid prairies of Texas (Bailey, 1980, 1998). Vegetation is primarily comprised of an oak woodland and grassland matrix (Wills, 2006). Topography consists of rolling hills with limestone outcrops, rocky soils, and caves typical of the Edwards Plateau (Kastning, 1983). Semi-ephemeral streams and pools fluctuate throughout the year, usually peaking during the wettest month of May. Camp Bullis reports that >140 guzzlers are scattered throughout the property as catchments of rainwater for wildlife, although not all of these guzzlers are maintained and hold water.

### 2.2 | Cameras

Remote cameras (Reconyx® PC900, Holmen, WI, USA) were mounted on trees overlooking the focal sampling sites (water bodies)



**FIGURE 1** Map of study area, Camp Bullis, Texas. Sampling locations are shown as colored circles (pond-yellow, stream-green, and wildlife guzzler-black)

where obvious sign of animal visitation had occurred (e.g., tracks, trails, or scat). Cameras were mounted  $\leq 10$  m from the water and were programmed to record motion-activated images. Upon motion, the cameras took three photos that were 30 s apart, followed by a quiet period of 15 min. The memory cards and batteries of cameras were refreshed once per month, and the camera positioning adjusted depending on water level. From the camera trap data located at each water body at the time of sampling, we recorded: the number of hours since the last pig visit to the water body; the number of images in the last day, week, and month with pigs; and the average group size of pigs in pictures with pigs using the Colorado Parks and Wildlife Photo Database (v3.0) for image processing (Ivan & Newkirk, 2016). These data were collected to help us assess our ability to successfully collect wild pig eDNA after a documented visitation to the site.

### 2.3 | Field eDNA collection

Previous studies some of us developed (Williams, Huyvaert, & Piaggio, 2016; Williams, Huyvaert, Vercauteren, Davis, & Piaggio, 2018; Williams et al., 2017) a method for detecting pig DNA in water. Typically, wild pigs use smaller bodies of water or edges of water

bodies. The activity of pigs and the nature of the water bodies they use mean that the target water bodies are often very turbid with high degree of suspended sediment and floating debris. Thus, we specifically tested methods for capturing DNA from such systems (Williams et al., 2017). Further, to meet our agency's goal of detecting wild pigs, we had some practical considerations to build into our method development: (a) our field personnel could not transport extra equipment such as filters, (b) field personnel did not have time to conduct filtering of samples in the field, (c) field samples could not require a cold chain for field preservation, and (d) samples collected had to be made small and light. Thus, the collection implemented in this study reflects these practical considerations as incorporated into the optimized collection and lab assay we developed for wild pig detection (Williams et al., 2016, 2017, 2018).

We selected 12 sites within Camp Bullis for water sample collection (Figure 1). We stratified the sites by three water body types (natural ponds, streams, and wildlife guzzlers). We randomly selected four sites from each of three water body types: natural ponds, stream (broken into 100 m segments), and wildlife guzzlers. Water samples were collected once a month for 2 months (September and October 2016). We focused on this time period to avoid camera damage (i.e., March–May) associated with spring flooding and because our access was restricted to some of the sites during hunting season (i.e., November–December).

Each sampling event consisted of taking 10 water samples from each of 12 sites. Sampling was evenly spaced in wildlife guzzlers, and spread around ponds and along streams. The goal of taking multiple samples was to get a diverse selection of water from the site to maximize the chance that the samples would contain wild pig DNA and overcome the heterogeneous distribution of eDNA in the environment (Furlan et al., 2016). Samples were collected by submerging a 60 ml Nalgene bottle ten centimeters below the surface of the water (when possible) until the bottle was filled to a line marked on the bottle at 45 ml. Then, 15 ml of Longmire's lysis buffer was added to the sample (1 part Longmire's: 3 parts sample water as in Williams et al., 2016). With the collection of each sample, our intent was to collect as little sediment as possible to avoid colloiddally bound DNA that may not have been shed recently (Barnes et al., 2014). A negative control (15 ml tap water carried by sampler with 5 ml of Longmires buffer added) was collected during each sampling session at each site. Gloves were worn at all times while sampling and were changed between sites. Collectors were instructed not to walk in the water body to avoid contamination between sites. Each bottle was labeled with a unique ID relating to site, field replicate number, and date collected. The locality information, number of samples, type of water body (i.e., wallow, moving, artificial waterer/tank, other), approximation of size of water body (i.e., small (<10 m<sup>2</sup>), medium (10–1,000 m<sup>2</sup>), large (>1,000 m<sup>2</sup>)), pH, approximation of depth where the sample was collected (cm), if it was collected along a transect or randomly, and whether there was evidence of pig activity in the area (i.e., tracks, rooting, wallowing) were all recorded as site-level characteristics. Samples were stored in a box at ambient temperature

until being shipped to the United State Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) National Wildlife Research Center (NWRC) within a week of collection. Once received at NWRC, the samples were placed in a –80°C freezer until further processing.

## 2.4 | eDNA capture, extraction, and amplification

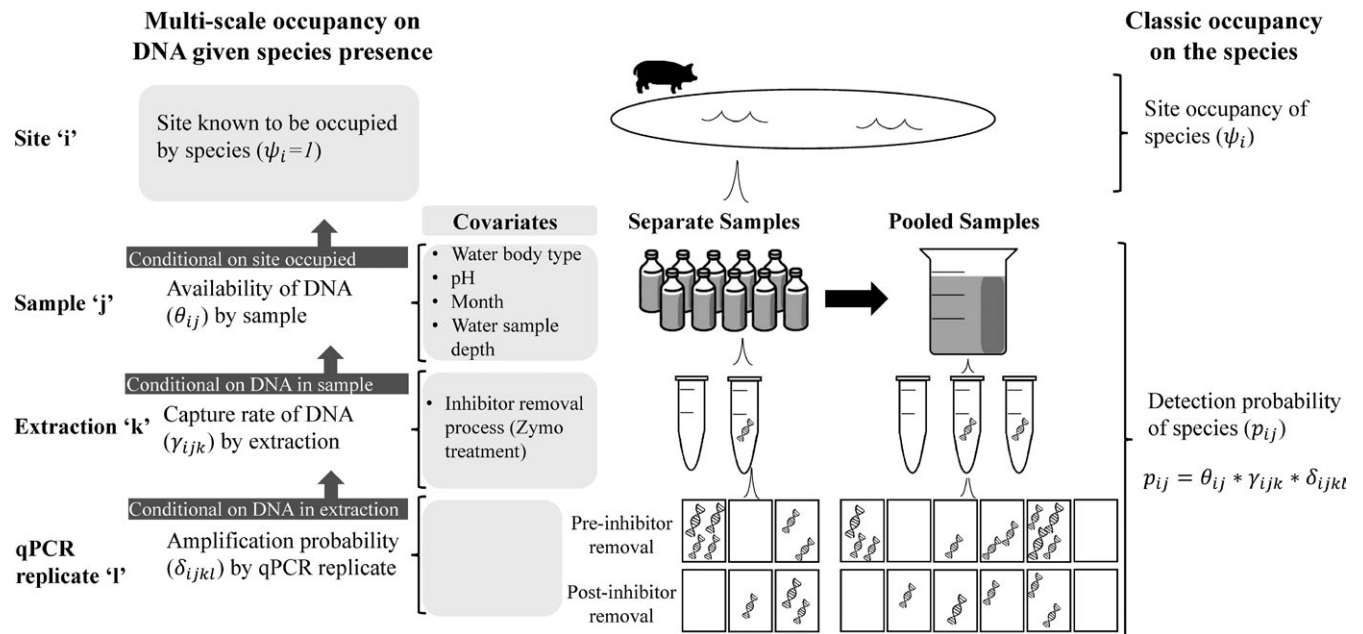
We compared two strategies for our extraction and amplification procedures. For the first procedure, we followed protocols that recommend pooling samples by site (Biggs et al., 2015) and conducting three extractions from the pooled sample (Piaggio et al., 2014). From each extraction, we ran five qPCR replicates before inhibitor removal and 5 qPCR replicates after inhibitor removal (Figure 2; details on inhibitor removal below). This method was more efficient for laboratory work; however, it produced few positive water samples despite camera data showing pigs at sampled water bodies. We suspected that pooling all 10 samples from each site diluted the DNA below detectable levels. Therefore, for our second strategy, we examined the 10 samples by site separately and conducted two extractions per sample. From each extraction, we ran three qPCR replicates before inhibitor removal and three qPCR replicates after inhibitor removal (Figure 2). Therefore, each individual sample was split in half, with one half being pooled by site and the second half being treated separately.

The analyses of the separate water samples by site increased the detection probability substantially (see results) but proved burdensome for laboratory personnel and resources. Therefore, for the second month (October), we compared the amplification for pooled samples as in the previous month, but we additionally used a removal sampling approach on the separate samples to reduce laboratory costs. For the removal sampling approach, we analyzed one water sample at a time (conducting two extractions per sample and three qPCR replicates per extraction), per site and stopped once a positive detection occurred. For each sample, we proceeded as for the standard sampling design in which we took two extractions per sample and ran six qPCR replicates per extraction (three pre-inhibitor removal and three post-inhibitor removal).

Each field sample for both months had a total volume of 60 ml (45 ml sample water + 15 ml Longmires = 60 ml). Therefore, 30 ml from each of the field samples were pooled and 30 ml were kept separate. Each extraction for both pooled and separate samples was conducted on 15 ml subsamples. We conducted three extractions of 15mls from our pooled water samples and two extractions of 15 ml for our separate samples.

We followed an optimized eDNA extraction protocol for detecting wild pig DNA in turbid waters (Williams et al., 2017). Briefly, this involved centrifuging each 15 ml subsample at 9,000 g for 15 min at room temperature, extracting DNA from the pellet with the DNeasy mericon Food Kit (Qiagen) in triplicate. We included a negative control in each set of extractions to monitor for contamination.

The number of qPCR replicates varied by month for pooled samples (first month of study: five replicates, second and third month



**FIGURE 2** Schematic showing the different levels of uncertainty. On the right hand side is the classic two-level occupancy model structure (MacKenzie et al., 2002) with the occupancy of the species at the site level ( $\psi_i$ ) and the overall detection probability by sample “j” at site “i” ( $p_{ij}$ ). On the left, the detection probability is split into different levels (similar to: Schmidt et al., 2013): availability of DNA by sample ( $\theta_{ij}$ ) given pigs are at the site, capture rate of DNA by extraction ( $\gamma_{ijk}$ ) given DNA was available in the sample, and amplification probability by qPCR replicate ( $\delta_{ijkl}$ ) given DNA was captured in the extraction. Covariates are shown by the level they are modeled on

of study: six replicates) and were conducted on a CFX96 (BioRad) following Williams et al. (2017). Each qPCR reaction was a 30  $\mu$ l reaction. Each reaction contained 15  $\mu$ l Taqman environmental mastermix (Life Technology), 1  $\mu$ l BSA, 6  $\mu$ l distilled water, 1  $\mu$ l of each primer (10  $\mu$ mol/L), 1  $\mu$ l of the probe (2.5  $\mu$ mol/L), and 5  $\mu$ l of template DNA. Amplifications were performed on a Biorad real-time PCR thermocycler (Biorad, Hercules, CA, USA). The real-time thermocycling was a single cycle for 10 min at 95°C, then 50 cycles of 95°C for 15 s, and a final extension of 1 min at 52°C. A standard curve was run in triplicate with each run using dilutions of a synthetic sequence of our target amplicon (gBlocks® Gene Fragments, IDT). We included three negative controls with each qPCR reaction and also attempted amplification from extraction and field-collected negative controls to monitor for contamination.

For analysis of each of the 10 samples per site (no pooling), we performed two extractions per sample (15 ml each) for September and October with three qPCR technical replicates per extraction. Further, each extraction, from both pooled and individual water samples, was amplified with qPCR before and after OneStep Inhibitor Removal kit (Zymo Research) to determine the influence of removing inhibitors. All instruments and bench tops were decontaminated after each run with 10% bleach and all steps associated with pre-PCR, PCR, and post-PCR were conducted in different laboratory rooms, each dedicated to the processing of low quality/quantity DNA. Our assay was sensitive enough to detect down to 1 copy/ $\mu$ l of DNA (LOD—limit of detection). DNA extracts may be heterogeneous in the distribution of DNA and since only 5  $\mu$ l out of the 150  $\mu$ l elution of extract is used, we may not have transferred enough

template DNA to be successfully amplified successfully. Inhibitors are humic substances that may be coextracted with DNA and interfere with downstream processing (PCR), and are therefore can affect the probability that PCR will be successful (Matheson, Gurney, Esau, & Lehto, 2010; McKee, Spear, & Pierson, 2015).

## 2.5 | Analytical methods

To estimate the presence of wild pig DNA at each level of sampling, we adopted the multi-scale occupancy framework developed by Nichols et al. (2008) and was first proposed for use with eDNA by Schmidt et al. (2013) to include multiple levels in the observation process to the classic occupancy model (Figure 2; MacKenzie et al., 2002, 2006). When using eDNA to detect a species, we considered the following levels of the observation process (Figure 2): (a) the sampling process level which describes the probability that DNA of the study species is available for detection at the sample level,  $j$ , given that the species is present at site  $i$  ( $\theta_{ij}$ ), (b) the capture process level which describes the probability DNA is captured at the extraction level,  $k$ , given DNA is in sample  $j$  ( $\gamma_{ijk}$ ), and (c) the amplification process level which describes probability of amplification in a qPCR assay,  $l$ , of the sample given DNA is captured in extraction  $k$  ( $\delta_{ijkl}$ ; Kendall & White, 2009; Schmidt et al., 2013). The multi-level occupancy model can be written as a series of Bernoulli random variables such that  $z_i$  represents the true presence/absence status of the species of interest at site  $i$ ;  $a_{ij}$  is the availability status of the DNA in sample  $j$  given the species is present at site  $i$ ;  $d_{ijk}$  is the capture status of DNA in extraction  $k$ , of sample  $j$ , at site  $i$ ; and  $y_{ijkl}$  is the amplification

status of the qPCR assay,  $l$ , given DNA is present at site  $i$ , available in sample  $j$ , and captured in extraction  $k$ .

$$z_i \sim \text{Bernoulli}(\psi_i) \quad (1)$$

$$a_{ij} | z_i \sim \text{Bernoulli}(z_i \theta_{ij}) \quad (2)$$

$$d_{ijk} | a_{ij} \sim \text{Bernoulli}(a_{ij} \gamma_{ijk}) \quad (3)$$

$$y_{ijkl} | d_{ijk} \sim \text{Bernoulli}(d_{ijk} \delta_{ijkl}) \quad (4)$$

Since we knew the true occupancy status at all of our sites was 1 (confirmed by camera data and pig sign at all sites), we fixed the top level ( $\psi_i = 1$ ) and conducted a three-level multi-scale occupancy analysis (Nichols et al., 2008) to parse out the variability associated at the levels of water samples, extractions, and qPCR replicates. Detection data include only detections ( $y = 1$ ) and non-detections ( $y = 0$ ). To estimate each parameter in the three level model, we need replication at each level in the hierarchy ( $\theta_{ij}$ ,  $\gamma_{ijk}$ ,  $\delta_{ijkl}$ ). We only had replicates at each level in the months where we kept the water samples separate, thus only data from months 2 and 3 were used in this analysis.

For the multi-scale analysis, we examined the potential relationship of DNA availability and the pH of the water, the water body type (i.e., wildlife guzzler, pond, and stream), and the water sample depth (all of these were site-level characteristics; Figure 2). Additionally, we allowed availability to vary by month. The only factor we examined on the capture rate by extraction was the inhibitor removal treatment. In addition, we had several covariates from the camera data that we thought might predict DNA availability by sample including: hours since last pig visited the water body, the number of pictures with pigs within the last month, and the average group size when pigs visited the water body. We examined DNA availability both with and without the camera data. The camera data were used to provide a better understanding of the biological processes that relate to DNA availability.

We used the three level multi-scale occupancy model to understand which factors influence the observation processes at different levels, satisfying our first objective. By better understanding the observation process, we should be better able to evaluate the biological state of interest (i.e., occupancy of the species). Our second objective was to determine whether we could accurately assess the status of animal presence in an area with known wild pig populations. Therefore, we used a standard two-level occupancy model (classic occupancy) where the detection parameter ( $p_{ij}$  here) is the product of the DNA availability probability by sample ( $\theta_{ij}$ ), the capture probability by extraction ( $\gamma_{ijk}$ ), and the DNA amplification probability ( $\delta_{ijkl}$ ); the species detection status is given by  $w_{ij}$ . We used the information gained regarding important covariates from the three-level, multi-scale analysis to inform the detection process in this analysis.

For the classical occupancy analysis, we condensed the detection data such that if any qPCR replicate in any extraction had a detection, the detection history would be a 1, otherwise, it would be a 0. We used the pre- and post- inhibitor removal treatment periods as our two detection occasions. We used this approach to compare occupancy estimates between pooled and separate samples and among months.

Our third objective was to evaluate the power to correctly detect animal presence under different conditions and provide guidance on eDNA sampling design that would be most reliable and efficient in the field and laboratory. We used estimates from the multi-scale occupancy analysis (where  $\psi = 1$ ). These provided us with realistic estimates of the probability of detecting wild pigs at different water body types, under different water conditions, and pre- and post- inhibitor removal treatment. We ignored information from the cameras as this type of data would not be available in most eDNA based field studies. To estimate the cumulative probability of detection (denoted with an  $x^*$ ), given sampling effort (number of samples/extractions/replicates,  $n$ ) for the availability of DNA by sample ( $\theta$ ), capture probability of DNA through extraction ( $\gamma$ ), and the amplification probability of qPCR ( $\delta$ ), we used equation 6. The variance for the cumulative probability was calculated using the delta method (equation 7; Powell, 2007). We used the cumulative probabilities to determine the minimum number of samples per site, extraction replicates, and qPCR replicates to achieve 90% detection probabilities at those respective levels. We used the cumulative probabilities of detection to determine the minimum number of samples per site, extraction replicates per sample, and qPCR replicates per extraction to achieve 90% detection probability for each level of the observation process/observation hierarchy. Using equations 5 and 6, we calculated the cumulative probability of detection given sampling effort for different ranges of effort (Supporting Information Appendix S1).

$$p_{ij} = \theta_{ij} * \gamma_{ijk} * \delta_{ijkl} \quad (5)$$

$$x^* = 1 - (1 - x)^n \quad (6)$$

$$\text{var}(x^*) = n^2 (1 - x)^{2*(n-1)} \text{var}(x) \quad (7)$$

We analyzed all data using occupancy models in Program MARK (Cooch & White, 2016; White & Burnham, 1999). We used the single-season, multi-scale occupancy option (Nichols et al., 2008) for the multi-scale analysis. We used the standard occupancy estimation method for the classic occupancy analysis. For both analyses, we treated the sites across months independently as we were interested in the independent assessments for each month to determine if there was a temporal component to the estimates of occupancy, and because the estimation of extinction and colonization parameters would be extraneous nuisance parameters. Model comparison was conducted using Akaike's Information Criterion corrected for small samples sizes (AICc, Burnham & Anderson, 2002; Doherty, White, & Burnham, 2012). For the multi-scale occupancy model investigating

relationships with availability of DNA and qPCR amplification success, we examined all additive combinations of covariates (including the inhibitor removal on capture rate and pH, month, water body type, water depth, pictures of pigs per month, average group size per picture, and hours since last visit on availability). The cumulative covariate weights, covariate weights  $>0.5$  were considered to be important (Doherty et al., 2012). When model uncertainty (multiple models within 2  $\Delta$ AICc of the top model) existed, we used model averaging to estimate occupancy and detection rates (Burnham & Anderson, 2002).

### 3 | RESULTS

In month one, one stream site was dry and was not sampled. All sites were sampled in month two. During the study, there were 12,842 photographs taken from the 12 cameras (one at each site). There were 1,003 photographs of wild pigs (264 at pond sites, 428 at stream sites, and 311 at wildlife guzzler sites; examples shown in Figure 3). For the laboratory analyses, we had both positive and negative controls in all reactions and at all steps and we did not get amplification in any of our negative controls.

The results from the multi-scale occupancy analysis (objective 1) showed that DNA availability at the individual sample level was influenced by the month in which the sampling was conducted and the pH and the type of water body (Table 1, Supporting Information Appendix S3). September had a higher availability of DNA than October ( $\beta = 1.31$ ,  $SE = 0.05$ ; Figure 4). The availability of DNA was higher in sites with pH values closer to 7 than 8 ( $\beta = -1.65$ ,  $SE = 0.52$ ; Figure 4). The average pH value in our study was 7.6 ( $SD = 0.47$ ; Supporting Information Appendix S2). DNA availability tended to be higher for stream samples than for ponds or wildlife guzzlers, however, the 95% confidence intervals overlapped ( $\beta_{\text{guzzler}} = -1.18$ ,  $SE = 0.60$ ;  $\beta_{\text{pond}} = -1.34$ ,  $SE = 0.55$ ; Figure 4). We examined a linear trend with pH but demonstrate the estimated availability associated with two values of pH to avoid extrapolation past of this trend outside the range of data we examined (Figure 4). The depth of the water sample was not strongly related to DNA availability (Table 1). Sample depths averaged 8.6 cm ( $SD = 2.8$ , Supporting Information Appendix S2). Confidence intervals for point estimates in figures rely on the asymptotic normality of maximum likelihood estimates and are back transformed from the logit link estimates and standard error.

When camera data were included, the month of sampling and pH was still important predictors of availability but water body type was not (Table 1, Supporting Information Appendix S4). Instead, the number of photographs with pigs within the last month was related to DNA availability ( $\beta = 0.09$ ,  $SE = 0.03$ ; Table 1) and was substantially higher when more photographs with wild pigs were observed within the last month (Figure 5). The average number of photographs with pigs in the last month was 25.5 ( $SD = 49.8$ , Supporting Information Appendix S2). Neither of the other covariates from the camera data was strongly related to DNA availability. The average



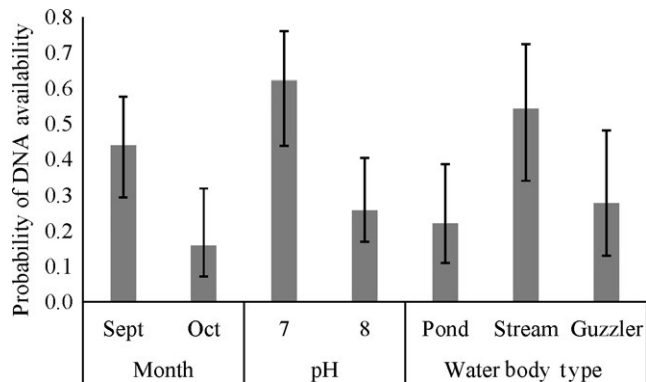
**FIGURE 3** Example images of wild pigs at pond sites (a), stream sites (b), and wildlife guzzler sites (c). Images are from camera traps located at eDNA collection sites on Camp Bullis, TX

number of hours since the last pig detection was 319 ( $SD = 434$ , Supporting Information Appendix S2), and the average group size in photographs was 1.48 ( $SD = 0.37$ , Supporting Information Appendix S2).



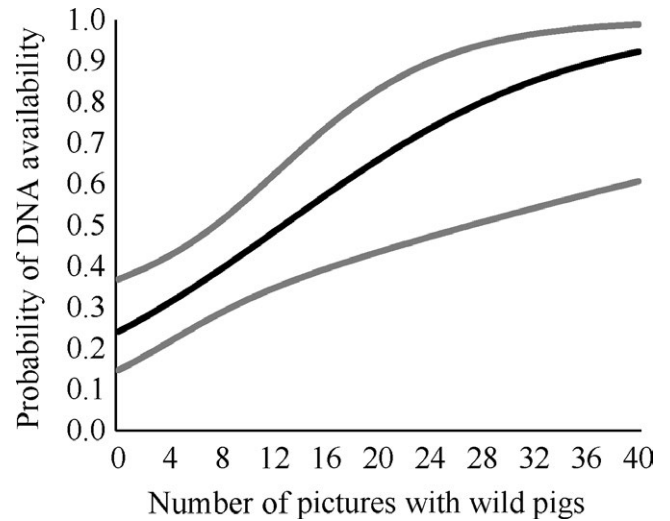
**TABLE 1** Cumulative covariate weights from multi-scale occupancy model selection procedure comparing relative covariate relationships between DNA availability at the sample level ( $\theta$ ) and capture rate at the extraction level ( $\gamma$ ) from eDNA samples at Camp Bullis, TX Sept–Oct 2016. Covariates are separated by parameter (availability or detection). Covariate weights are shown for models including and excluding camera data

Covariate by parameter	Cumulative covariate weight	
	Excluding camera data	Including camera data
Availability ( $\theta$ )		
pH	0.98	0.98
Month	0.92	0.97
Pictures per month	–	0.95
Group size	–	0.42
Water body type	0.80	0.23
Hours since last visit	–	0.30
Water sample depth	0.35	0.28
Capture rate ( $\gamma$ )		
Inhibitor removal	0.31	0.31



**FIGURE 4** Relationships of covariates with the probability of DNA availability at the sample level ( $\theta_{ij}$ ). Average probability of availability from the model containing important covariates (top model) by month with 95% confidence interval. The relationship between pH of the water body and the DNA availability with 95% confidence intervals. The relationship between water body type and eDNA availability with 95% CIs. Each estimate per covariate (month, pH, and water body type) is given as an average across other covariates

The impact of the inhibitor removal treatment did not have a substantial effect on capture rate by extraction ( $\beta = -0.16$  SE 0.34). The estimated capture rate without treatment was 0.49 (SE 0.07), whereas the estimated capture rate with treatment was 0.43 (SE 0.07). However, there were several instances where wild pig DNA was only captured without the inhibitor removal treatment and several instances where wild pig DNA was only captured from extractions with the inhibitor removal treatment. Therefore, even



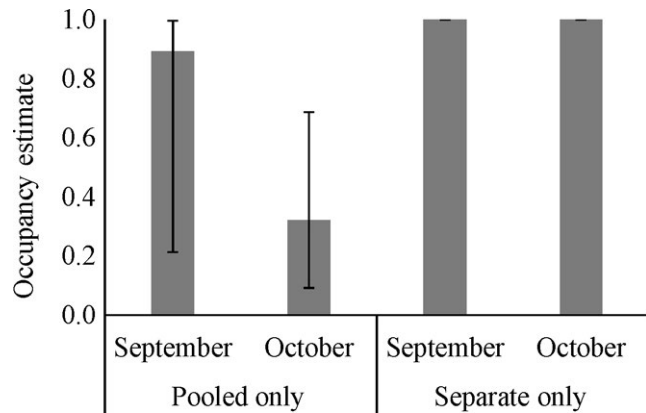
**FIGURE 5** The relationship between DNA availability and the number of pictures with pigs in the month prior to sampling and the DNA availability with 95% confidence intervals

though capture rates were similar either before or after the inhibitor removal treatment, the combined capture rates including both captures before and after the inhibitor removal did increase the overall apparent detection (raw samples without treatment that had a detection = 0.6; raw samples with treatment had a detection = 0.5, raw samples with or without treatment that had a detection = 0.79). No covariates were examined with reference to the qPCR amplification rate. The estimated amplification rate was 0.38 (SE 0.04).

To determine if we could correctly detect pig presence (objective 2), we used the important covariates from the previous analysis as covariates on detection ( $p$ ) to estimate overall occupancy (excluding camera data as they would not be present in a true field study). When only the pooled data were considered, the occupancy estimates were biased low for all months, but the estimates were highest in September (Figure 6). Occupancy was correctly estimated to be 1 when separate samples per site we examined for both the standard sampling design (all 10 water samples were analyzed, month 2) and for the removal sampling design (processing samples ceased once a detection was found by sample, month 3; Figure 6). For comparison, we also examined how the estimates in September may have changed had we used a removal sampling approach and not the full data set. We found that occupancy was estimated to be the same as when the standard sampling design was used, but the detection probability was lower ( $p = 0.22$ , SE = 0.09 for removal sampling compared to  $p = 0.29$ , SE = 0.07 for standard sampling).

We used the probabilities of DNA availability under each of the conditions (month, pH, water body type), while holding the other values constant at their mean, to examine the cumulative probability of availability under different numbers of water samples. To achieve a mean 90% cumulative availability probability in September, we only would have needed five separate samples (95% CI: 3–8 samples), whereas in October we would have needed 15 (95% CI: 5–25+ samples; Figure 7, using equations 6 and 7). Water pH had a strong influence the number

of water samples as when pH is 7 we would need five samples (95% CI: 3–7 samples) to ensure at least one detection but if pH was 8 we would need 18 samples (95% CI: 6–25+ samples) per site. The influence of water body type demonstrated that wildlife guzzlers and ponds required 9 (95% CI: 4–14 samples) and 10 (95% CI: 5–15 samples) samples per site, respectively, but streams would only need four samples (95%



**FIGURE 6** Occupancy estimates (with 95% confidence intervals) of wild pigs from eDNA samples at sites at Camp Bullis, TX Sept–Oct 2016. Estimates by month are shown for pooled versus separate samples

CI: 2–5 samples) for 90% cumulative probability of availability. The cumulative capture probability at the extraction level was slightly higher without inhibitor removal than with this treatment, requiring only four extractions per sample (95% CI: 3–5) compared to five (95% CI: 3–6) to achieve a 90% cumulative capture probability. Conducting qPCR amplifications on both types of samples (with and without inhibitor removal) showed a substantial improvement on DNA capture rates (Figure 7) requiring only two extractions to achieve 90% cumulative probability (95% CI 2–3 samples). Five qPCR replicates per extraction were required to achieve a 90% cumulative amplification probability.

## 4 | DISCUSSION

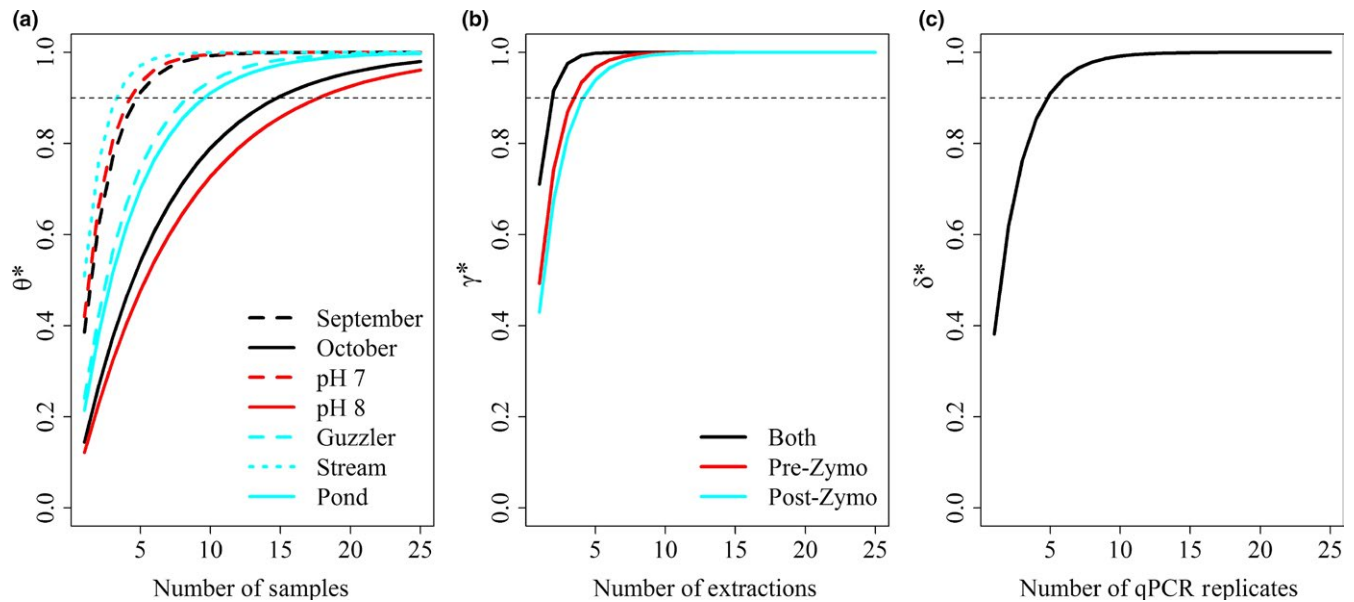
Understanding the observation process is an important first step for quantifying species distribution processes, especially for practical purposes in low-density populations (i.e., early detection of invasive species or timely detection of endangered species). When detection does not occur, it does not confirm that the species is absent—there is the possibility that the species is present but that detection probability was below 100% due to variability in observation processes (Schmidt et al., 2013). Our results showed that for observation processes that span multiple levels of uncertainty (e.g., non-invasive detection methods such as eDNA), it is important to understand the role of different factors across these levels and which environmental factors may have the strongest influence on ecological inference. Dissecting the observation process not only allowed for better inference of ecological processes but also provided a platform for

planning sampling adaptively, in a way that minimizes laboratory costs and time while maximizing species detection probability.

We expected differences in detection probabilities of wild pig eDNA ( $p$ ) among water body types due to variations in pig behavior and visitation rates, and abiotic conditions that affect DNA. Detection rates were similar between wildlife guzzlers and ponds, which might retain DNA more similarly when compared to streams. We expected streams to have a lower detection rate assuming that the movement of water would relate to a lower retention rate of DNA. However, streams had the highest detection rates of all water bodies in our study while ponds and guzzlers had similarly lower detection rates. Our study was conducted between the seasonal end of the summer and the beginning of the fall, when water levels were lower and thus streams were intermittent with pooling and low flow. These areas of multiple pools along a streambed proved conducive for DNA retention and detection, perhaps due to their attractiveness to pigs for multiple wallowing sites in proximity to one another (suggested by photographic evidence).

The probability of detecting DNA is likely affected by water quality (Barnes & Turner, 2016; Barnes et al., 2014). If sampling was conducted when water resources were plentiful, detection probability would have been lower as the DNA would have been more diluted. The probability of availability of DNA was highest in September when temperatures in the area were higher with less rainfall (water was scarcer) than in October when detection rates were lower. If sampling were conducted during cooler times of year, when there is more standing water, the detection rates would likely have been lower than we observed. Lower detection rates would mean that more field samples would need to be collected, and thus when designing an eDNA monitoring method seasonally varying detection rates will be important to incorporate in practice.

Consistent with previous studies, we found pH to be a strong indicator of DNA availability in water samples. Lorenz and Wackernagel (1987) found that DNA has a higher rate of adsorption to sand, and thus is more available, when the pH is neutral (pH 7). Additionally, DNA is known to degrade more rapidly or adsorb to certain soil particle types when pH deviates from neutral (Barnes et al., 2014; Lorenz & Wackernagel, 1987). In particular, Strickler, Fremier, and Goldberg (2015) and Seymour et al. (2018) found that degradation rates were higher in more acidic conditions, and degradation rates were lower or similar in alkaline compared to neutral conditions (respectively). Interestingly, we found higher detection rates in neutral conditions compared to slightly alkaline conditions. The fact that pH had such a strong relationship with DNA availability was surprising given the range we observed was narrow (7–8.4). Considerably, more samples would need to be collected to ensure a cumulative availability rate >90% when the pH of the water body was close to 8 compared to 7 (18 samples would be required compared to 5). In addition, pH may also be correlated with other factors we did not measure that influence wild pig behavior (e.g., plant community composition, turbidity of the water). Regardless, it appears that pH is an important factor in the probability of detecting eDNA and the variability in observation processes using this method. Thus, if pH is not appropriately



**FIGURE 7** Cumulative probabilities given a hypothetical number of replicates for three levels of the multi-scale analysis on detection using eDNA. The cumulative detection probability is the product of the different levels of uncertainty (Supporting Information Appendix S1). The cumulative availability of DNA ( $\theta^*$ , the probability that DNA is observed in at least one sample given the site is occupied) by water sample (a) is shown by the number of separate water samples collected based on estimates in September (dashed black line) and October (solid black line), for water pH values of 7 (dashed red line) and 8 (solid red line), and sampling from wildlife guzzlers (dashed blue line), streams (dotted blue line), and ponds (solid blue line). (b) The cumulative capture probability ( $\gamma^*$ , the probability that DNA is observed in at least one extraction given it is available in the sample) of DNA by number of extractions is shown for estimates based on pre-inhibitor removal treatment (solid red line), post-inhibitor removal treatment (solid blue line), and both pre- and post- inhibitor removal treatment (solid black line). (c) The cumulative amplification probability ( $\delta^*$ , probability that DNA is observed in at least one qPCR replicate given it is captured in the extraction) is shown by the number of qPCR replicates (solid black line). An interactive version can be found in the supplemental information

accounted for across geographic sites and seasons, species distributions based on eDNA could be biased.

We found the observation processes (detection probability) of wild pig from eDNA to be variable during our study. One possible explanation for that variability is that the behavior of wild pigs in the area may influence the detection probability. We used cameras to examine how detection related to pig behavior. It was surprising that the time period since camera detection was not strongly related to DNA availability given that wild pig eDNA degrades with time (Williams et al., 2018). Despite the fact that degradation rates are higher for a single wild pig relative to a group (Williams et al., 2018), group size also did not strongly correlate with DNA availability. However, the number of pictures of wild pigs visiting our sampled water sites was strongly related to detection probability, suggesting that the general level of wild pig activity influenced detection. Thus, higher level of use may act similarly to larger group sizes in terms of DNA degradation rates. Since we are particularly interested in detecting pigs at low densities using eDNA, our actual detection probability may be lower than calculated in this study when implemented for detections of wild pigs in newly invaded areas or areas where control has reduced the population to low densities. Therefore, larger number of samples should be taken in the field to offset the lower levels of DNA availability that may occur.

Replicates that do not provide the same detection (some are positive, and some are negative) may arise from qPCR instrumentation variability when at the lower limits of DNA quantity (Hunter et al., 2017). The assay we applied for detection of wild pig eDNA has a limit of detection of 10 copies/ $\mu$ l (>95% of 8 qPCR replicates of our standard curve amplified; Bustin et al., 2009). Thus, one level of variability in our observation process involves the detection probability of low quantity/quality DNA. We accounted for this by including amplification probability with qPCR in our model and considering any positive as a detection (assuming no false positives). This is a risk if the specificity of our test has the possibility for false positives. We tested for false positives due to cross-reactivity and found high specificity in our assay for wild pigs (Williams et al., 2017). Issues with specificity would lead to incorrectly declaring species present when it is absent (Type I error) and issues with sensitivity would lead to declaring a species absent when it is present (Type II error). Generally, with early detection for an invasive species, the risk of declaring a non-detection when the species is present (low detection probability or poor sensitivity) is a considerably worse type of error than poor specificity due to the potential damage that may occur if an invasion went undetected.

We were particularly focused on the issues of false negatives (which may be very costly in our case), and we were confident that false positives were not an issue in our study. However, there are

some cases where the cost of responding to a new invasion may be considerable and thus there would be substantial concern for declaring a species is present when it is not (false positive). To that end, some studies require that two or more PCR replicates confirm a detection before they will be declared a positive (Kriger, Hero, & Ashton, 2006). If false positives are possible at the PCR level, then modifications must be made to the occupancy analysis as that is a clear violation of the occupancy assumptions (Mackenzie, 2005). There are several advances to occupancy modeling that will accommodate false positives (Miller et al., 2011; Royle & Link, 2006) and approaches that have been incorporated into eDNA analyses (Ficetola, Taberlet, & Coissac, 2016; Lahoz-Monfort, Guillera-Aroita, & Tingley, 2016).

Pooling field-collected eDNA samples by site and taking extractions from subsamples of the pooled sample (Biggs et al., 2015; Piaggio et al., 2014) reduces laboratory costs over running multiple samples per site separately. Replicate samples per site are collected because the DNA distribution in water is heterogeneous (Furlan et al., 2016), such that some samples may not contain DNA. Pooling the field samples eliminates the ability to have true replicates to estimate the detection process. The subsampling of the pooled samples only allows for pseudo-replicates. Our results demonstrated that pooling field replicates greatly reduced the detection probability and thus added to the variability in observation processes when compared to analyzing the replicate samples separately. This suggests that although DNA may be present in the pooled samples, the act of pooling may increase the effect of dilution (perhaps requiring extra filtration) and the use of pseudo-replicates reduces our power to make inference on the hidden biological state of interest. Therefore, if pooled samples are used, a species may go undetected. For threatened or endangered species, this may result in failure to provide conservation measures, whereas for an invasive species a non-detection could lead to a new invasion progressing unchecked and allowing it to become established.

Although keeping samples per site separate greatly increased our detection probability, it also greatly increased the laboratory time and expenses. Instead of three extractions for each site, we conducted 20 extractions per site (2 per each of 10 field water samples). This method is likely prohibitively laborious and costly for use as a standard sampling procedure to detect invasive terrestrial mammals. Therefore, we adapted our approach to a removal design (by sample), which has been found to be the most efficient design for occupancy studies (MacKenzie & Royle, 2005). However, MacKenzie and Royle (2005) point out that removal designs are less robust to violations of assumptions compared to a standard design. For example, occupancy should be constant during the sample frame (if a site is occupied it should be occupied during the entire sampling period), detection rates will be biased low if violations of this assumption occur, but the bias will be greater for the removal design than the standard design. This suggests that the maximum number of samples needed for a removal design might be more than the total number of samples for a standard design. Based on our estimated detection probability of 0.2 and our high occupancy rate, MacKenzie

and Royle (2005) would recommend 23 samples be collected per site. Fortunately, the field costs of collecting additional water samples per site are marginal and extractions would only need to be conducted on additional samples if no detections were made from each previous extraction.

Our field collection protocols were strategically designed to streamline and simplify sampling so they could be used by a variety of agencies in a variety of field conditions (Williams et al., 2016). Our methods are species-specific—we focused on small water bodies which wild pigs are likely to use for drinking and wallowing to optimize detection of this terrestrial species. To come up with a generic method to sample these small, turbid water bodies, we needed to use smaller water samples than are often used for detecting aquatic species (e.g., Furlan et al., 2016; Pilliod et al., 2013; Spear et al., 2015). The smaller water volumes used in our study may contribute to the lower levels of detection that we observed compared to other studies. Our results suggest that the probability of detection would be increased by collecting more water samples per site (which would increase the overall volume of water collected per site). By collecting more samples, we will still be able to use the same protocol for sample collection (which helps with consistency across study areas) while addressing the issue that small water volume may have on detection probability.

The field costs to collect additional samples may be low, but processing those samples in the laboratory is not. The removal approach may reduce overall costs. However, if no detections occur then the costs will be the same as the standard sampling approach if all samples are analyzed. Although costs in this case may be high, eDNA has been found to be a cost-effective option for monitoring populations of aquatic species compared to trapping (Lugg, Griffiths, van Rooyen, Weeks, & Tingley, 2018) and may be found to be worth the costs when compared to potential costs of undetected invasions of invasive species.

Pooling samples per site may be used as a first step in a removal sampling approach, and if there is no detection, testing the field samples could proceed discretely and in succession. However, using the data from September, if we had used the pooled data as a first step in a removal sampling approach (3 extractions and 12 qPCR replicates) and then continued with the separate samples (two extractions and six qPCR replicates) until we came to a detection, we would have needed a total of 228 qPCR replicates to analyze all of the data. If we had forgone the pooled samples and just conducted the removal approach on the separate field samples, we would have needed only 84 qPCR replicates to analyze all the data. Thus to detect rare individuals, we recommend a removal approach, with eDNA analysis on the separate samples, to optimize detection probabilities and reduce false negatives and variability in observation processes. A removal approach is also advantageous as collecting field samples are relatively inexpensive relative to the requisite cost of associated lab work; so this method reduces costs while not compromising power.

Using the detection probabilities of wild pig eDNA ( $p$ ) estimated from the separate samples per site, we calculated the cumulative detection probability given the number of samples analyzed through

qPCR to evaluate cumulative detection rates of using more samples than we collected, similar to Schmidt et al. (2013). Because availability varied by month, pH, and water body type, the number of samples necessary per site to have a high cumulative availability probability ( $\theta^*$ , where  $\theta$  is plugged into equation 8, and “n” is the number of samples) was dependent on site conditions. For example, in a stream sampled in September, we had over a 90% cumulative availability probability with only three samples analyzed, but in streams in October we needed to analyze at least six samples to have the same cumulative availability probability. While in wildlife guzzlers or ponds in September, we needed to analyze at least five samples for the same level of confidence and in October, we did not reach a 90% cumulative availability probability until 16 samples were analyzed. Thus by collecting a potentially superfluous number of water samples in the field, future studies will be better equipped to handle issues of poor DNA availability that may occur due to study timing, water body type, or other environmental conditions that reduce overall detection probability.

We conducted qPCR both pre- and post- inhibitor removal treatment. The treatment strips away inhibitors that prevent qPCR from amplifying DNA. However, in cases where there is a low quantity of DNA in the sample, the inhibitor removal treatment may result in reducing the DNA and not just the inhibitors. This might suggest that it is the number of extractions that is important and not the inhibitor removal process. However, we found many instances of detections that only happened pre- treatment and some that only happened post-treatment. Additionally, DNA concentrations of extracts before inhibitor removal treatment were low. Although, conducting one extraction with data both pre- and post- inhibitor removal resulted in a similar cumulative DNA capture rate to conducting two extractions with data either just pre-inhibitor removal or just post-inhibitor removal, we recommend that all samples are subject to qPCR amplifications both with and without inhibitor removal to maximize detection probability of target DNA or that an internal positive control (IPC) be incorporated into the qPCR assay to monitor inhibition (Goldberg et al., 2016).

## 5 | CONCLUSIONS

We found that observation processes of an invasive terrestrial mammal using eDNA are variable and dependent on the conditions of the water body sampled and laboratory processes. We recommend collecting a minimum of 10 water samples per site, but 20 samples would be better, and using a removal approach to laboratory analysis; more samples will be needed when the pH of the water body is not neutral. The adaptive analysis process will maximize detection while making efficient use of resources. We recommend running qPCR both pre- and post- inhibitor removal to increase capture rate of DNA. Variation in time of year and water body type will also impact the overall detection probability using eDNA and must be accounted for or estimates of species presence will be biased which could have serious implications for conservation or invasive species management. As has become part of best practices for non-invasive genetics studies (another approach for utilizing low quality/quantity DNA available in the absence of the target species), we

strongly advocate for conducting field studies on the system of interest to determine the factors influencing the study site and the target species (Lonsinger et al., 2015; Taberlet, Waits, & Luikart, 1999; Waits & Paetkau, 2005). Using multi-scale occupancy modeling for inference of species distributions will provide more robust inferences about factors affecting the observation processes of eDNA, which is particularly important for designing efficient sampling protocols. The multi-scale analysis may not need to be continued once greater understanding of the different levels of uncertainty are reached; however, new programs and packages (Dorazio & Erickson, 2017; Hunter et al., 2015; White & Burnham, 1999) have been developed that make performing this type of analysis marginally more work than a standard occupancy approach and therefore may be worth continuing. Improving the understanding of the observational process will help provide better understanding of the ecological processes influencing the distribution of the target species and guidance for future research and management.

## ACKNOWLEDGMENTS

We thank R. Tabor from Joint Base San Antonio, Camp Bullis for permission to conduct this research. C. Kohler (USDA/APHIS/NWRC) assisted in collecting environmental water samples. J. Halseth and E. Newkirk provided assistance with the photo database. Mention of commercial products or companies does not represent an endorsement by the US government.

## AUTHOR CONTRIBUTIONS

AJD, KEW, NPS, AJP, and KMP designed the research. KEW and NPS performed the research. AJD analyzed the data. AJD, KEW, NPS, AJP, and KMP wrote the paper.

## DATA ACCESSIBILITY

Detection data and related covariates are archived in Dryad data repository (<https://doi.org/10.5061/dryad.1fp259b>).

## ORCID

Amy J. Davis  <http://orcid.org/0000-0002-4962-9753>

Antoinette J. Piaggio  <http://orcid.org/0000-0002-4701-0746>

## REFERENCES

- Anderson, A., Sloatmaker, C., Harper, E., Holderieath, J., & Shwiff, S. A. (2016). Economic estimates of feral swine damage and control in 11 US states. *Crop Protection*, 89, 89–94. <https://doi.org/10.1016/j.cropro.2016.06.023>
- Bailey, R. G. (1980). Description of the ecoregions of the United States. US Department of Agriculture, Forest Service.
- Bailey, R. G. (1998). *Ecoregions, the ecosystem geography of the oceans and continents* (2nd ed.). New York, NY: Springer-Verlag.
- Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17, 1–17. <https://doi.org/10.1007/s10592-015-0775-4>

- Barnes, M. A., Turner, C. R., Jerde, C. L., Renshaw, M. A., Chadderton, W. L., & Lodge, D. M. (2014). Environmental conditions influence eDNA persistence in aquatic systems. *Environmental Science & Technology*, 48, 1819–1827. <https://doi.org/10.1021/es404734p>
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22, 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R. A., ... Dunn, F. (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation*, 183, 19–28. <https://doi.org/10.1016/j.biocon.2014.11.029>
- Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., ... de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29, 358–367. <https://doi.org/10.1016/j.tree.2014.04.003>
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference - A Practical Information-Theoretic Approach* (2nd ed.). New York, NY: Springer-Verlag.
- Bustin, S. A., Benes, V., Garson, J. A., Hellems, J., Huggett, J., Kubista, M., ... Wittwer, C. T. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55, 611–622. <https://doi.org/10.1373/clinchem.2008.112797>
- Chavarria, P. M., Lopez, R. R., Bowser, G., & Silvy, N. J. (2007). A landscape-level survey of feral hog impacts to natural resources of the Big Thicket National Preserve.
- Christy, M. T., Yackel Adams, A. A., Rodda, G. H., Savidge, J. A., & Tyrrell, C. L. (2010). Modelling detection probabilities to evaluate management and control tools for an invasive species. *Journal of Applied Ecology*, 47, 106–113. <https://doi.org/10.1111/j.1365-2664.2009.01753.x>
- Cooch, E. G., & White, G. C. (2016). Program mark: A gentle introduction. Retrieved from <http://www.phidot.org>
- Diefenbach, D. R., Marshall, M. R., Mattice, J. A., & Brauning, D. W. (2007). Incorporating availability for detection in estimates of bird abundance. *The Auk*, 124, 96–106.
- Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4, 675–683. <https://doi.org/10.1111/2041-210X.12060>
- Doherty, P. F., White, G. C., & Burnham, K. P. (2012). Comparison of model building and selection strategies. *Journal of Ornithology*, 152, 317–323. <https://doi.org/10.1007/s10336-010-0598-5>
- Doi, H., Katano, I., Sakata, Y., Souma, R., Kosuge, T., Nagano, M., ... Tojo, K. (2017). Detection of an endangered aquatic heteropteran using environmental DNA in a wetland ecosystem. *Royal Society Open Science*, 4, 170568. <https://doi.org/10.1098/rsos.170568>
- Dorazio, R. M., & Erickson, R. A. (2017). ednaoccupancy: An R package for multiscale occupancy modelling of environmental DNA data. *Molecular Ecology Resources*, 18, 368–380.
- Farnsworth, G. L., Pollock, K. H., Nichols, J. D., Simons, T. R., Hines, J. E., & Sauer, J. R. (2002). A removal model for estimating detection probabilities from point-count surveys. *The Auk*, 119, 414–425.
- Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources*, 16, 604–607. <https://doi.org/10.1111/1755-0998.12508>
- Furlan, E. M., Gleeson, D., Hardy, C. M., & Duncan, R. P. (2016). A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources*, 16, 641–654. <https://doi.org/10.1111/1755-0998.12483>
- Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., ... Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7, 1299–1307. <https://doi.org/10.1111/2041-210X.12595>
- Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, 116, 195–203. [https://doi.org/10.1016/S0006-3207\(03\)00190-3](https://doi.org/10.1016/S0006-3207(03)00190-3)
- Hone, J. (1983). A short-term evaluation of feral pig eradication at Willandra in western New South Wales. *Wildlife Research*, 10, 269–275. <https://doi.org/10.1071/WR9830269>
- Hunter, M. E., Dorazio, R. M., Butterfield, J. S. S., Meigs-Friend, G., Nico, L. G., & Ferrante, J. A. (2017). Detection limits of quantitative and digital PCR assays and their influence in presence-absence surveys of environmental DNA. *Molecular Ecology Resources*, 17, 221–229. <https://doi.org/10.1111/1755-0998.12619>
- Hunter, M. E., Oyler-McCance, S. J., Dorazio, R. M., Fike, J. A., Smith, B. J., Hunter, C. T., ... Hart, K. M. (2015). Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive Burmese pythons. *PLoS ONE*, 10, e0121655. <https://doi.org/10.1371/journal.pone.0121655>
- Ivan, J. S., & Newkirk, E. S. (2016). Cpw Photo Warehouse: A custom database to facilitate archiving, identifying, summarizing and managing photo data collected from camera traps. *Methods in Ecology and Evolution*, 7, 499–504. <https://doi.org/10.1111/2041-210X.12503>
- Kastning, E. H. (1983). Relict caves as evidence of landscape and aquifer evolution in a deeply dissected carbonate terrain: Southwest Edwards Plateau, Texas, USA. *Journal of Hydrology*, 61, 89–112. [https://doi.org/10.1016/0022-1694\(83\)90237-8](https://doi.org/10.1016/0022-1694(83)90237-8)
- Kendall, W. L., & White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, 46, 1182–1188.
- Korn, T., & Bomford, M. (1996). *Managing vertebrate pests: Feral pigs*. Canberra, ACT, Australia: Australian Government Publishing Service.
- Kruger, K. M., Hero, J.-M., & Ashton, K. J. (2006). Cost efficiency in the detection of chytridiomycosis using PCR assay. *Diseases of Aquatic Organisms*, 71, 149–154. <https://doi.org/10.3354/dao071149>
- Lahoz-Monfort, J. J., Guillera-Aroita, G., & Tingley, R. (2016). Statistical approaches to account for false-positive errors in environmental DNA samples. *Molecular Ecology Resources*, 16, 673–685. <https://doi.org/10.1111/1755-0998.12486>
- Lonsinger, R. C., Gese, E. M., Dempsey, S. J., Kluever, B. M., Johnson, T. R., & Waits, L. P. (2015). Balancing sample accumulation and DNA degradation rates to optimize noninvasive genetic sampling of sympatric carnivores. *Molecular Ecology Resources*, 15, 831–842. <https://doi.org/10.1111/1755-0998.12356>
- Lorenz, M. G., & Wackernagel, W. (1987). Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. *Applied and Environmental Microbiology*, 53, 2948–2952.
- Lugg, W. H., Griffiths, J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2018). Optimal survey designs for environmental DNA sampling. *Methods in Ecology and Evolution*, 9, 1049–1059. <https://doi.org/10.1111/2041-210X.12951>
- Mackenzie, D. I. (2005). Was it there? Dealing with imperfect detection for species presence/absence data. *Australian & New Zealand Journal of Statistics*, 47, 65–74. <https://doi.org/10.1111/j.1467-842X.2005.00372.x>
- MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248–2255.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy estimation and modeling – inferring patterns and dynamics of species occurrence*. Amsterdam, Netherlands: Elsevier.
- MacKenzie, D. I., & Royle, J. A. (2005). Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology*, 42, 1105–1114. <https://doi.org/10.1111/j.1365-2664.2005.01098.x>
- Matheson, C. D., Gurney, C., Esau, N., & Lehto, R. (2010). Assessing PCR inhibition from humic substances. *The Open Enzyme Inhibition Journal*, 3, 46–52.

- McClintock, B. T., Nichols, J. D., Bailey, L. L., MacKenzie, D. I., Kendall, W. L., & Franklin, A. B. (2010). Seeking a second opinion: Uncertainty in disease ecology. *Ecology Letters*, *13*, 659–674. <https://doi.org/10.1111/j.1461-0248.2010.01472.x>
- McKee, A. M., Spear, S. F., & Pierson, T. W. (2015). The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation*, *183*, 70–76. <https://doi.org/10.1016/j.biocon.2014.11.031>
- Miller, D. A., Nichols, J. D., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, *92*, 1422–1428. <https://doi.org/10.1890/10-1396.1>
- Nichols, J. D., Bailey, L. L., O'Connell, A. F. Jr, Talancy, N. W., Campbell Grant, E. H., Gilbert, A. T., ... Hines, J. E. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, *45*, 1321–1329. <https://doi.org/10.1111/j.1365-2664.2008.01509.x>
- Nichols, J. D., Hines, J. E., Sauer, J. R., Fallon, J. E., & Heglund, P. J. (2000). A double-observer approach for estimating detection probability and abundance from point counts. *The Auk*, *117*, 393–408.
- Piaggio, A. J., Engeman, R. M., Hopken, M. W., Humphrey, J. S., Keacher, K. L., Bruce, W. E., & Avery, M. L. (2014). Detecting an elusive invasive species: A diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Molecular Ecology Resources*, *14*, 374–380. <https://doi.org/10.1111/1755-0998.12180>
- Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*, 1123–1130. <https://doi.org/10.1139/cjfas-2013-0047>
- Powell, L. A. (2007). Approximating variance of demographic parameters using the Delta Method: A reference for avian biologists. *The Condor*, *109*, 949–954.
- Rodgers, T. W., & Mock, K. E. (2015). Drinking water as a source of environmental DNA for the detection of terrestrial wildlife species. *Conservation Genetics Resources*, *7*, 693–696. <https://doi.org/10.1007/s12686-015-0478-7>
- Rogala, J. K., Hebblewhite, M., Whittington, J., White, C. A., Coleshill, J., & Musiani, M. (2011). Human activity differentially redistributes large mammals in the Canadian Rockies National Parks. *Ecology and Society*, *16*, 16.
- Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, *87*, 835–841.
- Saunders, G., & Bryant, H. (1988). The evaluation of a feral pig eradication program during a simulated exotic disease outbreak. *Wildlife Research*, *15*, 73–81. <https://doi.org/10.1071/WR9880073>
- Schmelzle, M. C., & Kinziger, A. P. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources*, *16*, 895–908. <https://doi.org/10.1111/1755-0998.12501>
- Schmidt, B. R., Kéry, M., Ursenbacher, S., Hyman, O. J., & Collins, J. P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, *4*, 646–653. <https://doi.org/10.1111/2041-210X.12052>
- Seymour, M., Durance, I., Cosby, B. J., Ransom-Jones, E., Deiner, K., Ormerod, S. J., ... Creer, S. (2018). Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms. *Communications Biology*, *1*, 4. <https://doi.org/10.1038/s42003-017-0005-3>
- Simons, T. R., Alldredge, M. W., Pollock, K. H., & Wettroth, J. M. (2007). Experimental analysis of the auditory detection process on avian point counts. *The Auk*, *124*, 986–999.
- Snow, N. P., Jarzyna, M. A., & VerCauteren, K. C. (2017). Interpreting and predicting the spread of invasive wild pigs. *Journal of Applied Ecology*, *54*, 2022–2032. <https://doi.org/10.1111/1365-2664.12866>
- Spear, S. F., Groves, J. D., Williams, L. A., & Waits, L. P. (2015). Using environmental DNA methods to improve detectability in a hellbender (*Cryptobranchus alleganiensis*) monitoring program. *Biological Conservation*, *183*, 38–45. <https://doi.org/10.1016/j.biocon.2014.11.016>
- Strickler, K. M., Fremier, A. K., & Goldberg, C. S. (2015). Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, *183*, 85–92. <https://doi.org/10.1016/j.biocon.2014.11.038>
- Taberlet, P., Waits, L. P., & Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. *Trends in Ecology & Evolution*, *14*, 323–327. [https://doi.org/10.1016/S0169-5347\(99\)01637-7](https://doi.org/10.1016/S0169-5347(99)01637-7)
- Takahara, T., Minamoto, T., & Doi, H. (2013). Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*, *8*, e56584. <https://doi.org/10.1371/journal.pone.0056584>
- Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, *21*, 2565–2573. <https://doi.org/10.1111/j.1365-294X.2011.05418.x>
- Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., ... Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*, *17*, e63–e75.
- Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., ... Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*, 929–942. <https://doi.org/10.1111/mec.13428>
- Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *The Journal of Wildlife Management*, *69*, 1419–1433.
- West, B. C., Cooper, A. L., & Armstrong, J. B. (2009). Managing wild pigs: A technical guide. *Human-Wildlife Interactions Monograph*, *1*, 1–55.
- White, G. C., & Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study*, *46*, S120–S139. <https://doi.org/10.1080/00063659909477239>
- Williams, K. E., Huyvaert, K. P., & Piaggio, A. J. (2016). No filters, no fridges: A method for preservation of water samples for eDNA analysis. *BMC Research Notes*, *9*, 298. <https://doi.org/10.1186/s13104-016-2104-5>
- Williams, K. E., Huyvaert, K. P., & Piaggio, A. J. (2017). Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLoS ONE*, *12*, e0179282. <https://doi.org/10.1371/journal.pone.0179282>
- Williams, K. E., Huyvaert, K. P., Vercauteren, K. C., Davis, A. J., & Piaggio, A. J. (2018). Detection and persistence of environmental DNA from an invasive, terrestrial mammal. *Ecology and Evolution*, *8*, 688–695. <https://doi.org/10.1002/ece3.3698>
- Willoughby, J. R., Wijayawardena, B. K., Sundaram, M., Swihart, R. K., & DeWoody, J. A. (2016). The importance of including imperfect detection models in eDNA experimental design. *Molecular Ecology Resources*, *16*, 837–844. <https://doi.org/10.1111/1755-0998.12531>
- Wills, F. H. (2006). Historic vegetation of Camp Bullis and Camp Stanley, southeastern Edwards Plateau, Texas. *Texas Journal of Science*, *58*, 219–230.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Davis AJ, Williams KE, Snow NP, Pepin KM, Piaggio AJ. Accounting for observation processes across multiple levels of uncertainty improves inference of species distributions and guides adaptive sampling of environmental DNA. *Ecol Evol*. 2018;8:10879–10892. <https://doi.org/10.1002/ece3.4552>