Chapter 3

Sampling design for the plains fishes of the Arkansas River basin

The warm water reaches of the Arkansas River basin contain seventeen species of native plains fishes, plus over twenty introduced species.[a] Five of the seventeen native fishes are currently listed as special status by the State of Colorado: the Endangered Plains Minnow (*Hybognathus placitus*), Southern Redbelly Dace (*Phoxinus erythrogaster*), and Suckermouth Minnow (*Phenacobius mirabilis*); the Threatened Arkansas Darter (*Etheostoma cragini*); and the Species of Special Concern Flathead Chub (*Platygobio gracilus*) (http://cpw.state.co.us/learn/Pages/SOC-ThreatenedEndangeredList.aspx). Understanding the distributions of native plains fish species is essential to promoting conservation and potential expansion of remaining populations.

Baseline distributions of native species were determined through the intensive sampling efforts of Nesler et al. (1999). During the field seasons of 1993–1996, they

[a]Two species, the Arkansas River Shiner (*Notropis girardi*) and the Speckled Chub (*Macrhybopsis aestivalis*), are extirpated and are not considered in this count. The basin also includes native trout species in the headwaters (Cutthroat Trout, *Oncorhynchus clarkii*) that are excluded from our analyses and design.

visited 1,711 sites of which 6% (103) of the sites were wet but did not have fish and 75.7% (1,295) were dry at the time of visit. From the data, they calculated species compositions and relative abundances of fishes for twelve hydrographic units within the basin and created maps of species detections. The work culminated in the five special status listings mentioned in the preceding paragraph.

Over the past decade, Colorado Parks and Wildlife (CPW) has been resampling the basin to further monitor the native fishes in the basin. From 2008–2015, CPW sampled 141 sites throughout the warm water reaches of the basin on 201 different occasions. Through this sampling effort, they detected sixteen of the seventeen species expected in the basin; the River Carpsucker (*Carpiodes carpio*) was not detected. Fig. 3.1 shows the sampling locations, and Appendix 3.A contains an exploratory data analysis of the sampling efforts. A description of the protocols associated with the sampling may be found in Broms et al. (In Prep).

This report provides the locations where CPW should focus their future efforts and the procedures used to select those sites. Sampling locations were generated through a hybrid sampling design, with half the future sites selected through a space-filling design and the other half selected through an optimal adaptive sampling design, a model-based approach. The details associated with the selection of sites are described below.

3.0.1 Space-filling design

A space-filling design is a geometric approach to site selection. It does not rely on a model and it leads to a balanced set of sampling locations, with the sites chosen based on the distances between them. Generalized random tessellation stratified sampling (GRTS, Stevens and Olsen 2004) is a popular choice to select sites for natural resource monitoring, and it is the method that we used for the Arkansas

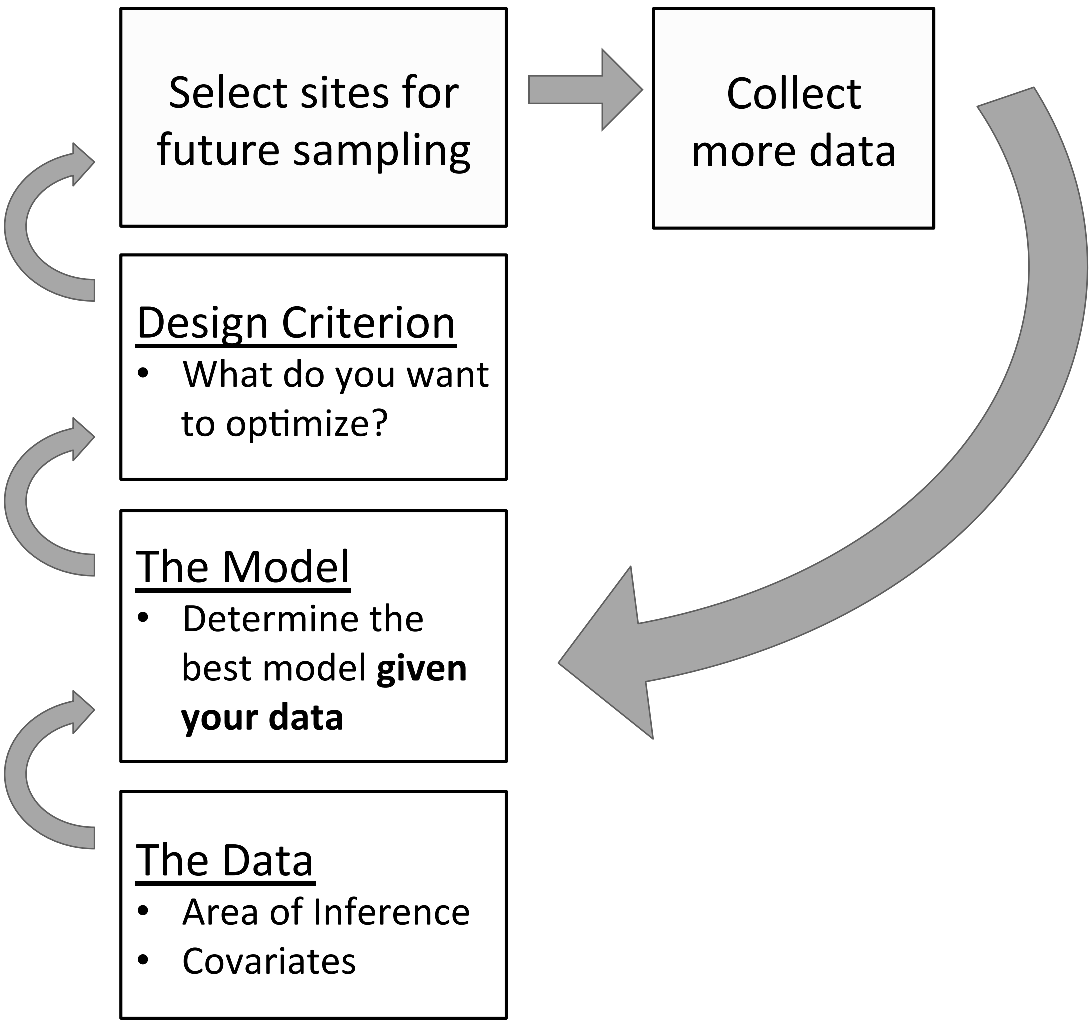
River basin to select the spatially balanced sites. The use of GRTS sampling allows for objective assessments of uncertainty and robust model estimates. Dobbie et al. (2008) provide a good overview of the approach for large-scale monitoring programs.

3.0.2 Optimal adaptive sampling design

Optimal adaptive sampling design is a model-based approach to site selection. Its use allows for improved uncertainty estimates over a design-based or geometric approach. For the Arkansas River basin, we used optimal adaptive sampling design to select half of the future sampling locations. The design is based on data collected since 2008 and directly focuses on reducing the variances associated with species distribution maps.

The site selection process involves fitting a model to data, selecting a design criterion statistic that quantifies how one would like to improve the inference, and running an algorithm to select the set of locations that optimizes the design criterion (Fig. 3.1). For the Arkansas River basin, we chose to minimize the variance associated with occupancy probability predictions. Because this method for site selection is more novel than GRTS, it is described in detail in the following sections of this report, and a more detailed description of optimal adaptive sampling design may be found in the South Platte chapter.

Figure 3.1: The building blocks of an optimal adaptive sampling design.



3.1 The data

Data are the foundation of the optimal adaptive sampling design. Organizing the data consisted of three components: cleaning the previously collected sampling data, determining the sampling frame or area of inference, and gathering the available covariates.

3.1.1 Previously collected sampling data

The first step was to gather and organize sampling data from the previous years of collection. The original data pull suggested 982 sampling occasions but this data set included repeated data, missing catch information, capture-mark-recapture studies, headwater sites, trapping data, and surveys conducted by USGS that followed a different protocol. After cleaning the data, 405 passes on 201 sampling occasions of 141 unique sites remained. A step-by-step description of the data organization process may be found in the '`ArkBasinDataCreation.docx`' file, and an exploratory analysis of the data may be found in Appendix 3.A.

As new data are collected, they should be added to the data files. The process of how to augment the current data with new data may be found in the '`ArkBasinDataCreation.docx`' file.

3.1.2 Area of Inference

Eastern Colorado is a semi-arid region with many ephemeral or intermittent streams and long-dry channel beds. Many of these stream flow lines may carry water and fish during and after flood events but not for the majority of year. Indeed, when Nesler et al. (1999) extensively sampled the basin in the 1990s, 76% of their sites were documented as dry. The NHD has a database of stream flow lines for the United States (USGS 2015), but many of the streams that they list as perennial in Colorado are in fact dry for most, if not all, of the year. Because it is inefficient for CPW to visit dry sites, a subset of the full stream network of the basin based on its likelihood of holding water was used as the sampling frame (Fig. 3.1). We recognize that additional habitat with ecological importance exists in the basin, but due to the logistical constraints of sampling, not all ecological valuable habitat exists within the sampling frame.

To create the sampling frame, we used ArcGIS 10.2. We started with the basin's flow lines from NHD, and excluded stream segments that were listed as 'NULL' or were above 1980m (6,496 feet) in order to exclude the cold, headwater streams of the basin. West of I-25, additional streams above 1830 meters (6,000 feet) were excluded based on expert opinion because the headwaters begin at lower elevations at these longitudes (*pers. comm. with P. Foutz*). We further limited the sampling frame to stream segments verified as holding water. Verification of a stream's perennial status was done through satellite imagery and personal communication with CPW biologists. Again, step-by-step details may be found in the '`ArkBasinDataCreation.docx`' file.

3.1.3 Covariates

We modeled fish distributions through multi-species occupancy models (Section 3.2, and MacKenzie et al. 2006). We gathered survey-level covariates taht may correlate with detections of fish and we gathered site-level covariates that may correlate with detections and with species distributions.. For the potential site-level covariates, we used longitude (`x`), latitude (`y`), elevation (`elev`), stream size (`size`), three land cover variables (`crops`, `dvlpd`, `wtlnds`), an indicator variable of whether or not a site was connected to the main stem of the Arkansas River (`unconnect`), a categorical variable indicating the reservoir that a site drained into (`reservoir`), and three additional indicator variables indicating the major stream that a site is associated with (`ftn`, `purg`, `main`; Table 3.1). The 100-meter resolution elevation (`elev`) was from the National Elevation Data set (http://ned.usgs.gov, accessed April 2014). For computational stability, `elev` was scaled by its mean and standard deviation so that it was centered at 0 with a standard deviation of 1. To obtain the stream size covariate (`size`), Strahler stream order principles were applied to our sampling frame. Stream size was used as a proxy for average stream flow. Land cover characteristics were obtained from the National Land Cover Database (Jin et al. 2013). To obtain the covariates: percent crop (`crop`), percent developed (`dvlpd`), and percent wetlands (`wtlnds`), we applied the ArcGIS Zonal Statistics 2 tool (ArcGIS 2015) to aggregate the land categorizations. The `crop` variable was calculated as the percent cropland within a 2000-meter buffer, the `dvlpd` variable was the percent of low, medium, or high urban density within a 1000-meter buffer, and the `wtlnds` variable was the percent of woody or emergent herbaceous wetlands within a 1000-meter buffer. A site was considered connected to the main stem of the Arkansas River if it was within 2,000 meters of other stream segments connected to the main stem (`unconnect`=0), otherwise it was considered unconnected (`unconnect`=1). Other local stream characteristics that are known or suspected to affect fish occurrence were not available throughout our study area and, thus, were excluded from our analyses.

We modeled detection probability as a function of the site-level covariates described above and the following survey-specific covariates (Table 3.2). Year (`year`) was included as a factor to account for changes in flow among years. Day-of-year (`yday`), scaled to be centered at 0 with a standard deviation of 1, and day-of-year squared ($\texttt{yday}^2$) modeled seasonal trends. We accounted for the removal sampling and multiple gear types by including pass number (`pass no`) and gear type (`seine`). Pass number (`pass no`) tested the influence of fish removal on subsequent passes. Method (`seine`) equaled 0 for an electrofishing pass and equaled 1 for a seining pass and allowed for different detection probabilities for the two gear types. Technician fatigue was included through the total count (`total ct`) variable, which was the logarithm of the total number of fish counted on the sampling occasion. We did not include the following survey-specific covariates because they were not available for all surveys: conductivity, time spent electro-shocking, seine mesh size, and number of seine hauls that constitute a pass.

Figure 3.1: The stream network of the Arkansas River basin used in our analyses and design. The stars and points represent the sites that were sampled. The black edging is the Colorado state boundary. The light gray polygon indicates the headwaters of the basin, and the dark grey polygon is our sampling frame. Note that the streams that exist in the headwaters are excluded from the map because they were excluded from our analyses.

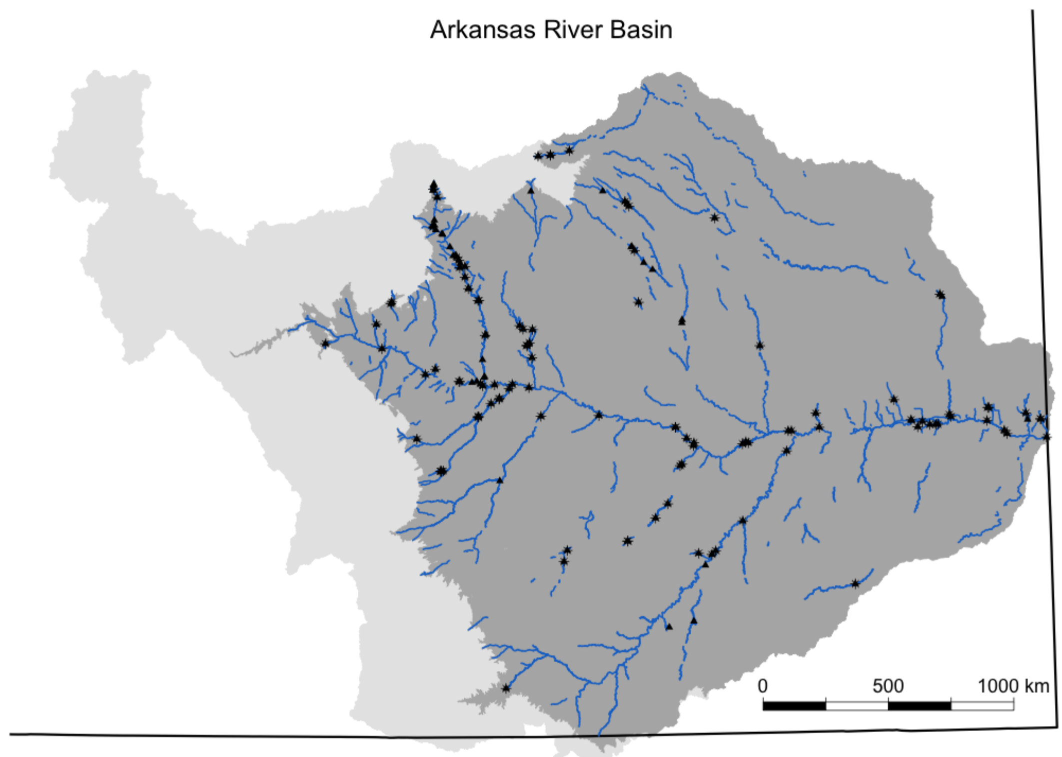


Table 3.1: Site-level covariates used in predicting the occupancy and detection probabilities for each species in the Arkansas River basin.

Covariate	Description
x	The latitude of the site scaled by the mean and standard deviation of all sites in the stream network.
x^2	The standardized latitude squared. Included to reflect the fact that a species' peak occurrences may occur in the middle of the stream network.
y	The longitude of the site scaled by the mean and standard deviation of all sites in the stream network.
y^2	The standardized longitude squared. Included to reflect the fact that a species' peak occurrences may occur in the middle of the stream network.
elev	A site's elevation scaled by the mean and standard deviation of all sites in the stream network.
$elev^2$	Elevation squared. Included to reflect the fact that a species' peak occurrences may occur in the middle of the stream network.
unconnect	A factor with two levels: Connected or Unconnected Stream. Connected is the reference level.
size	Stream size. In general, stream size is the Strahler Stream Order calculated solely from the streams included in the sampling frame
crops	Percent of land area within a 2000-meter buffer of a site that is cropland.
dvlpd	Percent of land area within a 1000-meter buffer of a site that is low, medium, or high density developed land.
wtlnds	Percent of land area within a 1000-meter buffer of a site that is wood or emergent herbaceous wetlands.
reservoirs	A categorical variable with three levels. A site may be above Pueblo Reservoir (abovePueblo), between Pueblo and John Martin reservoir (between), or below John Martin reservoir (belowJM). Between reservoirs is the reference group.
ftn	An indicator variable that equals 1 if a site is in Fountain Creek or drains into Fountain Creek, and equals 0 otherwise.
purg	An indicator variable that equals 1 if a site is on the Purgatoire River or drains into the Purgatoire River, and equals 0 otherwise.
main	An indicator variable that equals 1 if a site is on the main stem of the Arkansas River, and equals 0 otherwise.

Table 3.2: Survey-level covariates used in predicting the detection probabilities for each species in the Arkansas River basin.

Covariate	Description
`year`	Year as a factor so that detections change over time but without assuming a linear or quadratic trend. Data are from 2008-2015; year 2008 is the reference year.
`yday`	Day of year scaled by the mean and standard deviations from all surveys.
`yday`^{2}	The scaled day covariate squared.
`pass no`	The pass number associated with the given survey. An integer that runs from one to five.
`seine`	An indicator variable that equals zero if the survey was an electroshocking pass and equals 1 if it was a seining pass.
`total ct`	The logarithm of the total number of fish counted during a sampling occasion. The variable is included as a proxy for technician fatigue; if more fish must be processed, it may be that technicians are less likely to detect new species.

3.2 The model

We fit a Bayesian multi-species occupancy model with a known number of species to the data to gain inference on the occupancy status of multiple fish species:

$$y_{ijk} \sim \text{Bernoulli}(p_{ijk} \cdot z_{ij}) \tag{3.1}$$

$$\text{logit}(p_{ijk}) = \mathbf{v}'_{jk}\boldsymbol{\alpha}_i \tag{3.2}$$

$$z_{ij} \sim \text{Bernoulli}(\psi_{ij}) \tag{3.3}$$

$$\text{logit}(\psi_{ij}) = \mathbf{x}'_j\boldsymbol{\beta}_i \tag{3.4}$$

$$\boldsymbol{\alpha}_i \sim \text{N}\left(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha\right) \tag{3.5}$$

$$\boldsymbol{\beta}_i \sim \text{N}\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right) \tag{3.6}$$

$$\tag{3.7}$$

where y_{ijk} represent the detection of species i, $i = 1, \ldots, N$, at site j, $j = 1, \ldots, J$, on survey (visit) k, $k = 1, \ldots, K_j$; and K_j is the number of sampling occasions or surveys of site j (Broms et al. 2016). The observation y_{ijk} is 1 if species i occurred at site j and was detected on survey k, and is zero otherwise. The true occupancy of species i at site j, z_{ij}, is a latent indicator variable. If the species was detected at the site, then we know it occurred at the site and $z_{ij} = 1$; otherwise, its value is inferred from fitting the model. The probability of species i occurring at site j is ψ_{ij}, a function of site-level covariates. The probability of detecting the species is p_{ijk}, a function of site-level and survey-level covariates. The $\mathbf{v}_{jk}$ are the covariates pertaining to survey k of site j that affect detection probabilities and the $\mathbf{x}_j$ are the covariates pertaining to site j that affect occupancy probabilities. Note that the model assumes no false-positive detections of the species. For the multi-species occupancy model, the coefficients are random effects that depend on a community-level set of parameters ($\boldsymbol{\mu}_\alpha, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\alpha$, and $\boldsymbol{\Sigma}_\beta$). To complete the multi-species occupancy model, we used weakly informative priors

$$\boldsymbol{\mu}_\alpha \sim \mathrm{N}(\mathbf{0}, 2.25^2\mathbf{I}) \tag{3.8}$$

$$\Sigma_\alpha \equiv \sigma_\alpha^2\mathbf{I} \tag{3.9}$$

$$\sigma_\alpha^2 \sim \text{half-Cauchy}(2.25^2) \tag{3.10}$$

$$\boldsymbol{\mu}_\beta \sim \mathrm{N}(\mathbf{0}, 2.25^2\mathbf{I}) \tag{3.11}$$

$$\Sigma_\beta \equiv \sigma_\beta^2\mathbf{I} \tag{3.12}$$

$$\sigma_\beta^2 \sim \text{half-Cauchy}(2.25^2) \tag{3.13}$$

$$\tag{3.14}$$

For the Arkansas River, there were $N = 17$ native species and $J = 141$ unique sites. There was a maximum of $K_j = 17$ surveys at one site, but the median number of surveys per site was three. We fit over twenty models to the data with varying covariates to model the detection and occupancy probabilities (Table B1 in Appendix 3.B).

We fit the models using Markov Chain Monte Carlo (MCMC) as implemented in JAGS v3.4.0 (Plummer 2003). To choose a best-predicting model, we performed a five-fold cross-validation on over twenty models (Table B1) and calculated Logarithmic, AUC, and Brier scoring rules (Table 3.1). We folded the data at the site-level as in Broms et al. (2016). Formulas used to calculate the scores may be found in the Supplementary Materials associated with that article. For the cross-validation runs, we used 15,000 iterations with a burn-in of 5,000 iterations, and thinned by 10. We confirmed the convergence of the models with fewer iterations through visual examination of the trace, density, and autocorrelation plots. For the best-predicting model, as chosen through cross-validation and the logarithmic scoring rule, we fit the model using all the data. For this model, we obtained three chains with different starting values for each model for 30,000 iterations with a burn-in of 10,000 iterations and thinned by 10, leaving a total of 6,000 samples. We assessed convergence through visual examination of the trace, density, and autocorrelation plots. Goodness-of-fit of the model was checked through deviance residuals plots (Fig. 3.B.7 in Appendix 3.B, described in Broms et al. 2016).

The best-predicting model was the `simple size-2, no yr` model (Table 3.1). In this model, occupancy probabilities were a function of stream size. Detection probabilities were a function of method and stream size.

Note that the best-predicting model does not include any temporal covariates. If we kept `year` in by default, it makes a strong *a priori* assumption that we are optimizing the spatial design given information about year. However, there is not yet enough temporal data to base the design criterion on trends in time. The design criterion will optimize the design spatially using the previously collected data. It can be adapted each year, but it is not directly tied to dynamics in occupancy. The data may be used for other purposes down the road, after enough data are collected to learn about the dynamics of occupancy.

Predicted occupancy maps for three representative species of the basin, the Arkansas Darter (ARD), Longnose Dace (LND), and Suckermouth Minnow (SMM), and the undetected River Carpsucker (RCS) are presented in Fig. 3.1. Parameter values, estimated detection probabilities, and maps of predicted occupancy probabilities for all seventeen species of native fish may be found in Appendix 3.B.

Figure 3.1: Predicted occupancy probabilities for three representative species plus the species that went undetected during the study, the River Carpsucker (RCS). Scientific and common names for each species may be found in Appendix 3.A. Colored maps of the predicted occupancy and detection probabilities for all native species in the basin may be found in Appendix 3.B.

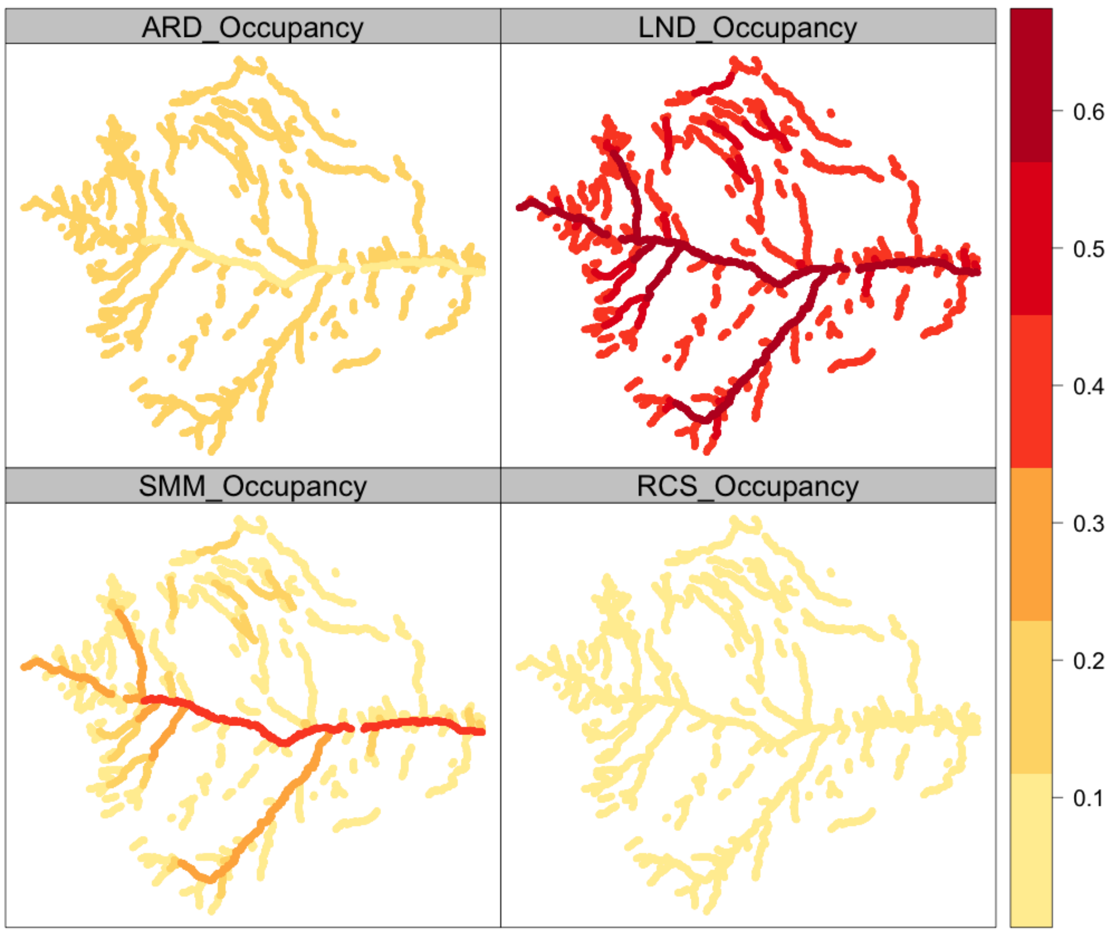


Table 3.1: Model selection statistics for the Arkansas River basin data. For all statistics except for Area Under the Curve (AUC), a lower value indicates a better-predicting model. For AUC, a higher value indicates a better-predicting model. The best-predicting model according to each statistic is highlighted in bold. The number of covariates, including the intercepts, associated with each model is represented by p. The Logarithmic, AUC, and Brier scores were calculated from the 5-fold cross-validation. The covariates that are included in each model may be found in Appendix 3.B.
In the best predicting model, occupancy probabilities were a function of stream size, and detection probabilities were a function of method and stream size.

Model	p	Logarithmic	AUC	Brier
`simpe size-2, no yr`	5	**631.0**	**0.724**	**183.1**
`simple size, no yr`	6	634.5	0.724	183.4
`simple size, count, no yr`	6	657.4	0.723	184.5
`simple alt stream-2, no yr`	8	668.7	0.723	187.7
`simple size-2`	12	679.0	0.717	187.0
`simple size`	14	707.8	0.715	188.1
`simple size, count`	13	717.2	0.720	187.6
`simple alt stream-2`	15	719.4	0.718	191.1
`simple land cover, size-2`	15	729.2	0.710	195.0
`simple alt stream`	17	750.7	0.714	193.1
`simple land cover, size`	18	757.6	0.709	196.3
`simple size, yday, no yr, connect`	7	769.6	0.703	196.6
`land cover, size`	20	816.8	0.692	205.2
`size`	16	830.6	0.691	208.7
`alt size`	18	831.9	0.693	210.8
`reservoir`	18	851.4	0.695	206.0
`combined-2`	21	895.1	0.693	212.4
`alt stream`	21	897.5	0.691	212.5
`no corrs`	21	906.5	0.690	213.3
`combined`	21	913.1	0.692	212.2
`no corrs-2`	20	923.5	0.687	217.5
`stream`	19	928.0	0.685	214.4
`full occu`	27	1011.7	0.673	234.3
`full occu-2`	29	1015.4	0.673	232.7
`full`	45	1364.5	0.668	247.7

3.3 The design criterion

The design criterion quantifies how we would like to improve our inference. Following the sampling design for the South Platte River basin, we chose to minimize the variances associated with the predicted occupancy probabilities,

$$q(d) = \sum_{m=1}^{M} \mathrm{var}(z_m|\mathbf{Y}, d), \tag{3.15}$$

where $m = 1, ..., M$ represent the potential sampling sites in the basin, z_m is the latent occurrence of a species at site m, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_J)'$ is a matrix of pre-existing data observations, and d is the proposed design. Our goal was to find a set of future sampling locations, d, that minimized this total variance. For the Arkansas River basin, the list of potential sites included 479 sites to represent the sampling frame plus the 141 previously sampled sites leading to a total of $M = 620$ potential sites. Eq. 3.15 is the theoretical formula for our design criterion for a single species. It is a function of the marginal posterior variance of $\mathbf{z}$ which is not directly available. Instead we approximated its value with the MCMC samples from our model-fitting.

Because z_m is a binary variable generated through a latent process in the Bayesian model, the full-conditional posterior distribution is Bernoulli and the variance associated with the variable is

$$\mathrm{var}(z_m|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{z}_{-m}, \mathbf{Y}, d) = \tilde{\psi}_m(1 - \tilde{\psi}_m), \tag{3.16}$$

where

$$\tilde{\psi}_m = \frac{\psi_m \prod_{j=1}^{J_m}(1 - p_{j,m})}{\psi_n \prod_{j=1}^{J_m}(1 - p_{j,m}) + 1 - \psi_m} \tag{3.17}$$

is the probability of a site being occupied given its sampling history. In the formula, ψ_m is the derived occupancy probability associated with site m, $p_{j,m}$ is the derived detection probability associated with survey j of site m, and J_m is the total number of surveys of site m (see Royle and Dorazio 2008 for derivation of Eq. 3.17). For the sites that were sampled, $J_m > 0$, and for the sites where no sampling occurs, $J_m = 0$ and $\tilde{\psi}_m = \psi_m$. The design criterion is now

$$q(d) = \frac{1}{S} \sum_{s=1}^{S} \sum_{m=1}^{M} \tilde{\psi}_n^{(s)} (1 - \tilde{\psi}_n^{(s)}). \tag{3.18}$$

where $s = 1, ..., S$ are the MCMC samples. We used the posterior mean as the design criterion. Similar superscripts should be added to Eq. 3.17 to signify that the occupancy and detection probabilities are also derived quantities in the Bayesian model:

$$\tilde{\psi}_m^{(s)} = \frac{\psi_m^{(s)} \prod_{j=1}^{J_m} (1 - p_{j,m}^{(s)})}{\psi_n^{(s)} \prod_{j=1}^{J_m} (1 - p_{j,m}^{(s)}) + 1 - \psi_m^{(s)}} \tag{3.19}$$

$$\psi_m^{(s)} = \text{logit}^{-1} \left(\mathbf{x}_j' \boldsymbol{\beta}_i^{(s)} \right) \tag{3.20}$$

$$p_{j,m}^{(s)} = \text{logit}^{-1} \left(\mathbf{v}_{jk}' \boldsymbol{\alpha}_i^{(s)} \right), \tag{3.21}$$

where the $\mathbf{v}_{jk}$ are the covariates pertaining to survey k of site j that affect detection probabilities and the $\mathbf{x}_j$ are the covariates pertaining to site j that affect occupancy probabilities. Notice that $q(d)$ depends only on the full-conditional posterior distributions and not the unknown response variables.

We calculated Eq. 3.18 for nine representative species of the basin: the Black Bullhead (BBH), Flathead Chub (FHC), Green Sunfish (SNF), Longnose Dace (LND), Orangespotted Sunfish (OSF), Plains Killifish (PKF), Red Shiner (RDS), Sand Shiner (SAH), and the Suckermouth Minnow (SMM), and averaged the design criterion over the nine species. Mathematically, our final design criterion was

$$q(d) \approx \frac{1}{9} \sum_{i=1}^{9} \frac{1}{S} \sum_{s=1}^{S} \sum_{m=1}^{M} \tilde{\psi}_{i,m}^{(s)} (1 - \tilde{\psi}_{i,m}^{(s)}) \tag{3.22}$$

where i indicates the nine species mentioned above. This equation gives each species equal weight but could be altered to weight some species more heavily or to change the species composition that $q(d)$ depends on.

Detection probabilities were functions of site-level covariates and survey-level covariates. In calculating the detection probabilities for future sampling efforts, we made assumptions related to the survey-level covariates so that each site would be considered equally. For each future sampling occasion, we assumed three surveys per site, two of which were electro-fishing passes and one seining pass, and we used the mean of the total count.

3.4 Future sampling locations

Following the procedures for the South Platte River basin design, we selected ten sites based on a space-filling design and ten sites based on the optimal adaptive sampling design for each year. Therefore, twenty sites were selected for each year: 2016, 2017, and 2018. The spatially-balanced sites from the space-filling design were selected first, since they are independent of the model, and then the optimal adaptive sampling design sites were chosen, assuming that the spatially balanced sites would be sampled.

To obtain the spatially-balanced sites, we implemented a generalized random tessellation stratified (GRTS) design with unequal sampling probabilities based on the stream size because of the geometry of the Arkansas River basin. The GRTS design ensured that the main stem of the Arkansas River, Fountain Creek, and Purgatoire River were adequately sampled, rather than allowing all future sampling to potentially be on low order streams. Streams of size 1 (intermittent and small streams) were given a weight of 0.40; streams of size 2 where given a weight of 0.10 because of the low number of such streams; streams of size 3, which included the Fountain Creek and Purgatoire River, were given a weight of 0.20; and streams of size 4, which represent the main stem of the Arkansas River, were given a weight of 0.30. We created a rotating panel design for three years. We based the design on a three-year rotation to mirror the design proposed for the South Platte and because the lifespan of many plains fishes is approximately three years. Different sites should be sampled every year for three years, and then the sites are repeated.

To select the optimal adaptive sampling design sites using the design criterion presented in Eq. 3.22, Section 3.3, we used a neighborhood exchange algorithm initiated through a simple search algorithm of 1,000 random designs. We used twenty neighbors per site as a compromise between a general exchange algorithm and the efficiency of less swapping. Because the exchange algorithm is guaranteed to converge to a local optima only, we repeated the process eight times to ensure that the lowest $q(d)$ was close to the global optima. Indeed, the same $q(d)$ value appeared for multiple starts, lending confidence that the $q(d)$ was the global optima.

For years 2017 and 2018, the same approach was used to select sites for sampling, but with the optimal sites selected assuming the previously selected sites were already incorporated into the data set. The random, optimal adaptive sampling design sites change every year and do not repeat.

Optimal adaptive sampling design results. If no additional sampling beyond the spatially balanced locations occurred, then $q(d) = 729.1$ (Fig. 3.1). If all potential sites were sampled, then $q(d) = 326.3$. With ten sites chosen through the algorithm described above, $q(d) = 718.5$ for 2016. Ten optimal adaptive sites were also chosen for 2017, and ten more for 2018. The final design lead to a design criterion value of $q(d) = 698.0$.

Fig. 3.2 displays the sites associated with the final design. Coordinates and other location details associated with the future sampling locations may be found in Appendix 3.C and in the '`ArkSamplingLocations.kml`' file.

Figure 3.1: The uncertainty in our occupancy predictions, represented through the design criterion, q, under different sampling schemes. Lower values are preferred. The horizontal, gray dotted line shows the design criterion value if only the space-filling sampling took place. The dotted line represents typical criterion values if simple random sampling was used to choose sites. The solid line are the criterion values when sites are optimally selected.

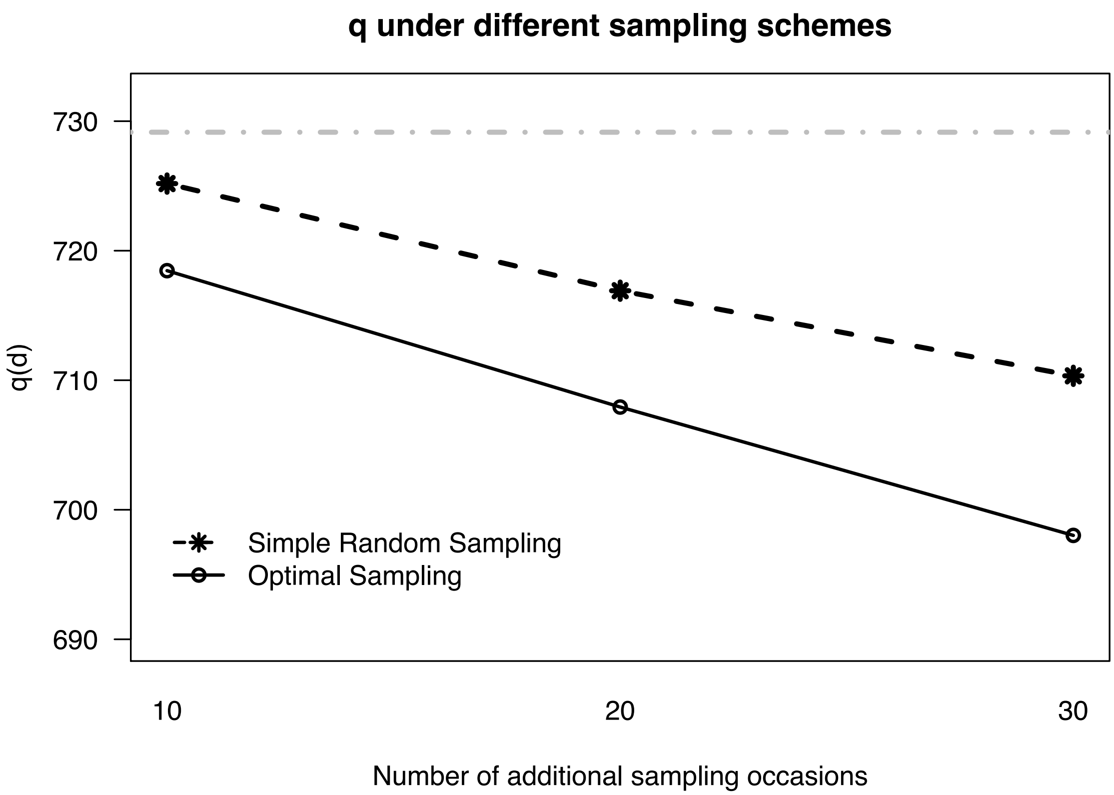


Figure 3.2: Sites to be sampled in 2016, 2017, and 2018. The black circles are the spatially balanced sites (also known as the GRTS or historic sites) that should be sampled each year. The cyan diamonds are the optimal adaptive sampling sites (also known as the OASD or random sites) that should be sampled each year. A list of the coordinates and water names associated with each site may be found in Appendix 3.C.