3.D Steps to select future sampling locations.

In order for the site selection to be reproducibile and to use the process for future years, implementation details are described here.

3.D.1 Data

The data was organized through the '`ArkBasinDataCreation.docx`' file, included in this appendix. The steps in this file describe how the sampling frame and the covariates were created.

To add additional sampling occasions to the data files, follow the steps at the end of the '`ArkBasinDataCreation.docx`' document. The data organization step culminates in an `ArkData.RData` R workspace to be used for the subsequent sampling design steps.

Ark. Basin Data Creation/Organization

This document describes how the raw data are cleaned and organized for the optimal adaptive sampling design. It contains three main sections. The first section creates the sampling frame and covariates. This section only needs to be revisited if the sampling frame changes and/or if additional covariates are needed. The second section adds the sampling occasions. This section needs to be revisited each year as more data are collected. The third section describes how the data are then organized in R. If any section is revisited, then subsequent sections will need to be rerun.

I. CREATE THE SAMPLING FRAME/ STREAM NETWORK MAP

Inputted files:

File Name	Description	File type and source
ua_annotated_map.shp	A set of flowlines that is often used by CPW to represent the Arkansas River basin.	Shapefile. From Grant.
NHDflowlines.shp	The set of potential waterways of the Arkansas River basin	Shapefile. Originally from Grant, but clipped to Ark Basin and with Nulll flowlines removed.
ArkBasinPolygon.shp	HUC4	Shapefile. Originally from Grant, but with only Ark Basin HUC selected.
ArkBasinBelow6000.shp	Polygon of the basin that exists solely below 6500 feet.	Shapefile. Derived from the ArkBasinPolygon.shp and elevation DEM files.
ArkBasinBelow6500	Polygon of the basin that exists solely below 6500 feet.	Shapefile. Derived from the ArkBasinPolygon.shp and elevation DEM files.
SampOccasions.dbf	Table of the locations of all Ark Basin sampling locations.	Created in R file: "OrgSampOccForExport.R" from database download from Andrew Treble.

[1] Also from: ftp://rockyftp.cr.usgs.gov/vdelivery/Datasets/Staged/Hydrography/NHD/State/

Outputted files:

File name	File type	Description
SurveyDesignStreams.shp	Shapefile	Polylines representing the sampling frame.

1. Begin with Grant's "ua_annotated_map" set of flowlines.
    a. Also load the NHDflowlines.shp, ArkBasinPolygon.shp, and ArkBasinBelow6000.shp files.
2. Clip "ua_annotated_map" streams to only include segments below 6496 ft (1980m). (Geoprocessing drop-down > Clip) Save as "SurveyStreamsTmp.shp".
3. Clip "NHDflowlines" to only include segments below 6496 ft (1980m).
4. Project streams into "NAD_1983_13N." (Toolbox > Data Management > Projections > Project)
5. Add the Sampling Locations to the map. **This step is done now to make sure all important waterways are include in map, but step may need to be repeated when more sampling is complete.** First, convert the "DATA_KristinBroms_RevisedPull_03172016.xlsx" file to a .csv file and then use R to convert the .csv file to a .dbf file using the "OrgSampOccForExport.R" file. Add the .dbf file to ArcGIS. Right click on the table, go to "Display XY Data" to display the data. Change the Symbology so that it is color-coded as to whether the site should be included or excluded. Export the table to create a .shp file of the data points. (Make sure the SamplingLocations.shp file has the same projection.)
6. Using the Editor Toolbar, remove from "SurveyStreamsTmp.shp" file:
    a. Ponds
    b. Lines that do not match any NHD flowlines
    c. Dry intermittent streams that do not match ARD assessment map.
7. Using the Editor Toolbar, add to "SurveyStreamsTmp.shp" file:
    a. NHD flowlines where sampling took place.
    b. Add wet NHD flowlines that look wet in Google Earth and were not in original map, when possible.
    c. Create ditches lines for sampling locations that do not exist on NHD flowlines.
    d. BE SURE TO CHANGE THE ATTRIBUTES ASSOCIATED WITH NEW FLOW LINES. ***
8. "Snap" sampling locations to stream segments. (And/or move some stream segments to match sampling locations—GIS line does not always match stream flow on Google Earth.) Toolbox > Editing > Snap. Use 100 Meters as distance.
    a. Double check that all sampling locations land on a stream segment, and add stream segments if necessary.
    b. Re-snap if necessary and/or use Editor toolbar to move points.
9. Confirm Stream Network and Sampling Locations (Exclude/Include) with P. Foutz.
10. The final sampling frame is the "SurveyDesignStreams.shp" shapefile.

B. ADD COVARIATES.- PART 1

Inputted files:

File name	Description	File type and source
SurveyDesignStreams.shp		Shapefile. Created above.
Elev100mNAD	100-meter elevations projected into NAD Zone 13N to match other data.	Raster. Originally from http//ned.usgs.gov, but projected into common projection.
Land Cover Covariates	Raster where each cell value represents a different land type.	Raster from National Land Cover Database.

Outputted files:

File name	Description	File type
SurveyDesignStreams.shp	Updated version of previously created file. Updated to include columns of covariates	Shapefile.
SurveyDesignStreamsAsPoints	Discretized version of the above file.	Shapefile.

1. Add Field to say if line is Perennial or (detached) Intermittent stream segment. A stream is intermittent if it is more than 2000m from the connected stream network. (Open Attribute Table. File > Add Field… Add a field named "Stream Type". Use selections (and ruler) to define if site is INT or PER.) Right click > Field Calculator. Put the string in double "quotations." **NOTE: In this step, we are using the term "perennial" to describe streams that are *connected* to the main stem of the Ark, and "intermittent to describe streams that are *unconnected*. In subsequent clean-up and R, the terminology is corrected.
2. Create categorical covariate: if a site is above Pueblo reservoir, between Pueblo and John Martin reservoir, or if a site is below John Martin reservoir. (Open Attribute Table. File > Add Field… Add a field named "ReservoirLocation". Use selections (and ruler) to define if site is "abovePueblo", "between", or "belowJM".)

3. Create an indicator covariate to separate Fountain Creek from rest of basin; include longitude to reflect changes along Fountain Creek. (Open Attribute Table. File > Add Field… Add a field named "FountainCrk". Use selections (and ruler) to define if site is YES or NO.)
4. Create an indicator covariate to separate Purgatoire River from rest of basin. (Open Attribute Table. File > Add Field… Add a field named "PURG". Use selections (and ruler) to define if site is YES or NO.)
5. Create an indicator covariate to separate the main stem of the Arkansas River from rest of basin. (Open Attribute Table. File > Add Field… Add a field named "MAIN". Use selections (and ruler) to define if site is YES or NO.)
6. Create Stream Size covariate. I calculated it by hand with the "Select" tool, Editing, and the "Field Calculator" but there is probably an easier way. (The existing field called STRAHLER does not make sense for all stream segments.) Open the attribute table of the "SurveyDesignStreams" file and Add a new Field-"Stream Size" as a Short Integer. Select by attributes, select all Intermittent streams and use the "Field Calculator" to set their Stream Size = 1. Then manually select other stream segments and create a stream size for them based on Strahler stream order properties. Change intermittent streams to correctly follow Strahler stream order properties as well.

7. Create zone covariate from Nesler et al. (1999) booklet, p. 53. NOT DONE. No map/ not enough in Nesler et al. (1999) for me to reproduce.

<u>C. ADD COVARIATES.- PART 2</u>

The continuous version of the stream network is represented by a series of stream segments. However, these stream segments are of very unequal length. To equalize every portion of the stream and make it more objective when calculating covariates, we will discretize it into a raster and then convert it to points.

1. Add relevant data: elevation raster, gradient raster (or create it, Spatial Analyst > Surface > Slope), and land cover feature class/shapefile. Make sure all have the "NAD_1983_13N" projection. If not, Toolbox > Data Management > Projections and Transformations > Raster > Project Raster.
2. Rasterize the stream network. (The steps below basically follow steps 6 and 9 in the "STARS: Spatial tools for the analysis of river systems version 2.0.0" Tutorial.) Each raster cell will then have a value equal to the reach ID. This field will help in joining stream names and other attributes to the stream points.
    a. Add the STARS toolbox.
    b. Make sure stream network ("SurveyDesignStreams" shapefile) has a projected coordinate system, e.g., "NAD_1983_13N".
    c. Enable Spatial Analyst Extension. Customize > Extensions. Make sure Spatial Analyst is checked.

d. Change the Environment Settings. Menu > Geoprocessing > Environments. Expand M Values and set "Output has M Values" to Disabled. Repeat for Z Values. Under Processing Extent, set Extent to Same as layer elev100mNAD. Under Raster Analysis, set cell size to Same as layer elev100mNAD. No Mask is used. Click OK.
e. Overwrite outputs by default. Geoprocessing > Geoprocessing Options > check box next to "Overwrite…"
f. Add a long integer field named reached to the stream attribute table. Calculate field: reachid = OBJECTID. (OBJECTID field could change.)
g. Convert the stream to raster format. Toolbox > Conversion> To Raster > Polyline to Raster. Use Value field= reachid. Zoom in and check that cells align with elevation.

3. Convert the stream raster into points. Toolbox > Converstion > From Raster > Raster to Point. Save the file as "SurveyDesignStreamsAsPoints".
   a. Add long integer field- reachid = grid_code. Used in R to join with continuous stream network and get stream names, those covariates, etc.
      i. Right click on "SurveyDesignStreamsAsPoints" and Join with "SurveyDesignStreams". Join by reachid. (Double-check how/if this changed files in ArcGIS vs doing it in R.

4. Add the elevation and gradient values to each stream point.
   a. Add long integer field Elevation, and double/float field Gradient.
   b. Spatial Analyst > Extraction > Extract Values to Points. Extract the elev100mNAD raster values to each point. This creates a new shapefile/feature class: "ArkStreamElevations."
   c. Spatial Analyst > Extraction > Extract Values to Points. Extract the grad100mNAD raster values to each point. This creates a new shapefile/feature class : "ArkStreamGradients." **NOTE: Because of misalignment between the elevation DEM and the NHD flowlines, the gradient values do not make sense and were not used in subsequent analyses.
   d. Add two new fields to the "SurveyDesignStreamsAsPoints.shp" file: Elevation and Gradient.
   e. Join the "SurveyDesignStreamsAsPoints.shp" file with the "ArkStreamElevations.shp". Join by pointid.
      i. After Join, Field Calculator… Elevation = RASTERVALUE.
      ii. Remove Join.
   f. Repeat with the "ArkStreamGradients" file.
      i. After Join, Field Calculator… gradient = RASTERVALUE.
      ii. Remove Join.

5. Create the land cover covariates.

a. First create buffered polygons from the stream points. Geoprocessing Drop-down > Buffer. Use a 1000m buffer, and then repeat with a 2000m buffer for the crops covariate. This will create two new shapefiles/ feature classes: "SurveyDesignStreamsAs1000mBuff" and "SurveyDesignStreamsAs2000mBuff," respectively.
b. The landcover raster has many different categories in it. We need to reclassify it into many different rasters with each land cover given its own raster. (This step mirrors step 10 of the STARS tutorial). First, make sure projections match the stream network. Spatial Analyst > Reclass > Reclassify. Set the Landscape cell value(s) of interest to 1, and all other cell values to 0. This step creates the cropsNAD.tiff, dvlpdNAD.tiff, and the wtlndsNAD.tiff raster files.
c. Now calculate the percent of the land area of each point that is covered by the given land type. Because these polygons overlap, we cannot use the "Zonal Statistics" tool and will instead use a "cheat".
    i. Convert the land cover raster to points. Toolbox > Conversion > From Raster > Raster to Points. Double-check the projection.
    ii. Create new Fields in the "SurveyDesignStreamsAsPoints" file to hold the crops, dvlpd, wtlnds covariates. I call the new fields "cropsCount", "dvlpdCount", "wtlndsCount".
    iii. Intersect the crop point file with the 2000m buffered polygon file. Use outpoint type = POINT. OK to create the files in a temporary directory here.
        1. Open the attribute table of the "CropsBuffIntersect" file. Right click the "FID_SurveyDesignStreamsAs2000mBuff" Field > Summarize… Create SumCrops table.
        2. Join the SumTable to the "SurveyDesignStreamsAsPoints" file (use the pointid Field). Some row will be <Null> because not all sites have crops nearby.
        3. Set "cropsCount" Field = BuffID Count to make permanent. (Click "yes" to the warning message.)
        4. Remove Join. Delete "CropsIntersect" file. Delete SummTable.
    iv. Repeat the above step with the dvlpd points file and the 1000m buffered polygon file.
    v. Repeat the above step with the wtlnds points file and the 1000m buffered polygon file.

D. EXPORT

1. The "SurveyDesignStreamsAsPoints" file should have (at least) the following fields/columns:
    i. reachid,
    ii. elevation,
    iii. gradient,

    iv. cropsCount,
    v. dvlpdCount,
    vi. wtlndsCount
2. The “SurveyDesignStreams” file should have (at least) the following fields/columns:
    i. GNIS_Name
    ii. reachid
    iii. Stream Type (INT or PER)
    iv. ReservoirLocation (abovePueblo, between, belowJM)
    v. FtnCreek (NO or YES)
    vi. PURG (NO or YES)
    vii. MAIN (NO or YES)
    viii. StreamSize (integer values 1 – 4)
3. Add UTM coords to the “SurveyDesignStreamsAsPoints” files. Add Long Integer fields. Then right-click > Calculate Geometry…
4. Export the two files as .dbf files (dBASE table). We will merge them in R through the “reachid” Field.

II. MATCHING SAMPLING POINTS TO STREAM-AS-POINTS NETWORK.

Inputted files:

File name	Description	File type and source
includeSampOcc.dbf**	This file is a subset of the SampOccasions.dbf file, which contains all potential sampling occasions. The file has been cleaned in R to exclude occasions that did not follow protocol.	Created in R file: “OrgSampOccForExport.R”
SurveyDesignStreamsAsPoints.shp		Shapefile. Created in previous steps.

** File must first be created in R.

Outputted files:

File name	Description	File type and source
snappedSurveys.dbf	A file very similar to the includeSampOcc.dbf file, but with new UTMs that match to points on the stream network.	Created through ArcGIS

1. First use the “Ark_OrgSampOccForExport.R” file to clean up the survey data (from the SampOccasions.xlsx file) and to create a file that only includes the sampling occasions to be used in the analyses (e.g., Arkansas R headwaters will be excluded).
    a. This file begins with the original data pull. The main spreadsheet of this file should have the following columns: SurveyPurpose, Protocol, Gear, TotalEffort, EffortMetric, WaterID, WaterName, SampleDate, Location, Elevation, UTMZone, UTMX, UTMY, Basin, StationLength, AvgWidth, StationCode, SiteType, SurveyID, SpeciesCode, Status, Catch, and Source.
    b. Add a Column called “NewStatus”, and a column called “sameStatus” right after the “Status” column. The Status codes were not uniform. For codes that did not follow protocol, e.g., only had numbers without a letter in front describing the sampling method, check the notes and see if you can fill in the correct status. The NewStatus column should then be similar to the original Status column, but with more information.
    c. Sometimes multiple passes were labelled: E1, E2, S1, S2, and sometimes the same four passes were labelled: E1, S2, S3, S4. Make the statuses consistent by using the former labelling. Therefore the “sameStatus” column should have the most consistent and updated via notes Status information.
2. Add the newly created “IncludedSampOcc.dbf” file to ArcGIS. And then right-click to Display Data… Double check that all sites that appear should be included and are near stream segments.
3. Toolbox > Editing > Snap. And snap the includeSampOcc Events to the “SurveyDesignStreamsAsPoints” file. I used a distance of 300 Meters. Once tool is completed, zoom in and make sure it worked and all sites are on the stream points. Remove files and repeat above 3 steps as necessary.
4. Open Attribute Table, and Export records as “snappedSurveys.dbf” (dBASE table). We will merge it with the “SurveyDesignStreamsAsPoints.dbf” file in R through their common UTMs.

<u>ADDING NEW SAMPLING OCCASIONS</u>

After each field season, add the new data to the SampOccasions.xlsx file. Repeat the above steps of the “MATCHING SAMPLING POINTS TO STREAM-AS-POINTS NETWORK.” section.

III. ORGANIZING DATA FOR MODELS AND DESIGN

Use the “Ark_OrganizeDataNew.R” file in R to create.

3.D.2 Model

Use the `RunCrossVal.R` file select a best-predicting model, and then use the `RunModels.R` file to fit the model to the full data set. You will need to have JAGS software installed on your computer. It is the output from the `RunModels.R` file that is used in the design criterion and site selection.

The `ModelResults.R` and `BestModResidPlots.R` files may be used to examine the output from the chosen model and to reproduce the figures and tables contained within this report.

3.D.3 Design criterion and selection of sites

First select the spatially balanced GRTS sites with the `SelectBalancedSites.R` file, and then select the optimal adaptive sampling sites with the `SelectOASDSites.R` file. A Google Earth file and .csv file to display and utilize the sampling locations may be created with the `ArkPlotFutureSites.R` file.