Data Pre-Processing R Script and GIS Standard Operating Procedures

This document describes how to use the R scripts created for the QA/QC process. These instructions assume the user has a basic understanding of how to use R and associated software. There are two different processes that are done with R. The first R script process downloads and formats the statewide POD flat file for use in the GIS Pre-Processing Template. Next, the GIS pre-processing workflow creates a subset of verified POD points that fall within an area of interest. Finally, the second R script process uses the output table generated by the GIS workflow to create the input files that will be copied into the excel calculator Modules.

If you have not already done so, download and unzip the Water Right QA/QC.zip file to your local hard drive.

Navigate to the project named QAQC_Scripts in the Pre Processing Scripts folder. When you click on the R project it will open automatically in RStudio.

> **If this step does not work reach out to the help desk so they can help download R and RStudio. R is the programing language and RStudio is the graphical user interface that R uses to make things easier for the user.

There are 4 main aspects of R studio that will be helpful to learn when running the scripts.
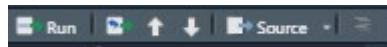
2. **Global Environment** – The Global Environment is where you can see and access your Data and Values. It can be found it in the top right window.
3. **Console** -The Console located in the bottom left of RStudio is the first spot you can see an error if one occurs. Most of the time the error is something small that can be fixed by reading the error.
4. **Files** – The file window on the bottom right will show where the files are being downloaded. You can also see output files from your local hard drive in each of the folders.

Steps to run the code

1) **install.packages** – Highlight and click run to install.packages.

```
install.packages("tidyverse")
install.packages("readxl")
install.packages("geosphere")
install.packages("stringi")
install.packages("stringr")
install.packages("pracma")
```

2) **Library** – Load the packages by highlighting and click run.

```
library(tidyverse)
library(readxl)
library(geosphere)
library(stringi)
library(stringr)
library(pracma)
```

3) **Memory** – If you do not have enough memory on your laptop/pc then you can artificially increase it for R. You do this by running the next line of code. You will know if this is necessary If the scripts do not run.

```
memory.limit()
memory.limit(size=56000)
```

4) **Preprocessing/GIS** – Run lines 19-48 and use this output file for the QAQC GIS Pre-Processing. It will be located in the output files.

5) **GIS** – Start by reviewing the GIS Pre-Process Overview document and open the POD Pre-Processing Template in Esri's ArcGIS Pro. Follow the instructions in the GIS Pre-Processing Procedures Part 1 document to incorporate the POD file created in step 4. Next, step through the GIS Pre-Processing Procedures Part 2 document and use ArcGIS Pro to clip and verify POD points in your Area of Interest. The GIS workflow produces a CSV file called APPID_MAX_MAF. Save this file to the pre-processing Input_Files folder for use in Step 6.

6) **Flat file filters** – Run lines 53-131 This step takes the list of application numbers, flat files, and then filters the flat files to a point it makes filtering easier for each process.

7) **Excel files for Modules** – Run lines 134-318. The files are located in the output folder in the file structure for the scripts. These will be used to copy and paste into the excel calculator Modules.

8) **Priority Date Module** – Copy and paste the data from step 7 into the Priority Date Module and process the data (see the Data Processing Module Procedures for more information on how to complete this step). The Priority Date Module must be completed first because the remaining modules require its results to formulate properly. The Priority Date Module output is then copied and pasted into the Input_file folder for the Missing RMS code in step 8.

8) **Missing RMS Reports** – Run lines 319-335. This will be the final output for the Missing RMS Report Module.

9) **QAQC Working File** – Run lines 336-355. This output is used for the QAQC working Module.

10) **Missing Contact Information** – Run lines 355-371. The output will give you the diverters that are missing contact information. (This step will not be possible for external use)