**Directory structure**

While there is some flexibility in the directory structure, it is strongly recommended to have the following directories in your base directory:

* **CalSim3\_Model\_Runs**
  + This directory mirrors the one on the google drive, located here:

*https://drive.google.com/drive/folders/1E67zovHLuPUiUtDtf-ju7R-ltCYG2DXH*

It needs to be downloaded to extract CalSim3 data from the DSS output files.

Note: only the subdirectory **Scenarios** needs to be downloaded at this level

* + - The **Scenarios** directory in turn contains a directory **Group\_Data\_Extraction**, and a number of directories containing CalSim3 studies/scenarios (e.g. **s0001\_DCR2023\_9.3.1\_danube\_hist**), as well as a number of Excel files containing a list of scenarios of interest and their attributes (e.g. **coeqwal\_cs3\_scenario\_listing\_v2.xlsx**, discussed below) and the definition of variables of interest and their attributes (e.g. **trend\_report\_variables\_v3.xlsx**, discussed below)
      * **Group\_Data\_Extraction** contains CSVs of grouped data extracted from a subset of the studies listed in **coeqwal\_cs3\_scenario\_listing\_v2.xlsx**, consisting of a subset of variables listed in **trend\_report\_variables\_v3.xlsx**. Note that the scenario identifier (e.g **s00001**) is appended to each variable name (e.g. **S\_SHSTA\_s0001**).
      * A Calsim3 scenario directory such as **s0001\_DCR2023\_9.3.1\_danube\_hist** contains two subdirectories, **Model\_Files**, which is the raw output of a Calsim3 run, and **Data\_Extraction**, which contains data extracted from the single scenario. In this case no scenario identifier is appended to variable names. Note that the data itself is contained in subdirectories of **Data\_Extraction**, indicating the variable listing file and tab used to generate the CSV, while the CSV file reflects the name of the source DSS (e.g. **Variables\_From\_trend\_report\_variables\_v3\_TrendReportVars\_CS3/** **DCR2023\_DV\_9.3.1\_Danube\_Hist\_v1.7.csv**)
  + Note that if only grouped data is desired, only the directory **Group\_Data\_Extraction** needs to be downloaded to the local machine. Similarly, if only data from individual studies is desired, then the various **Model\_Files** subdirectories need not be present on the local machine; these are only needed to extract new data from the CalSim3 runs.
* **coeqwal**
  + This directory is generated when the GitHub repository is downloaded from here:

*https://github.com/maramahmedd/coeqwal*

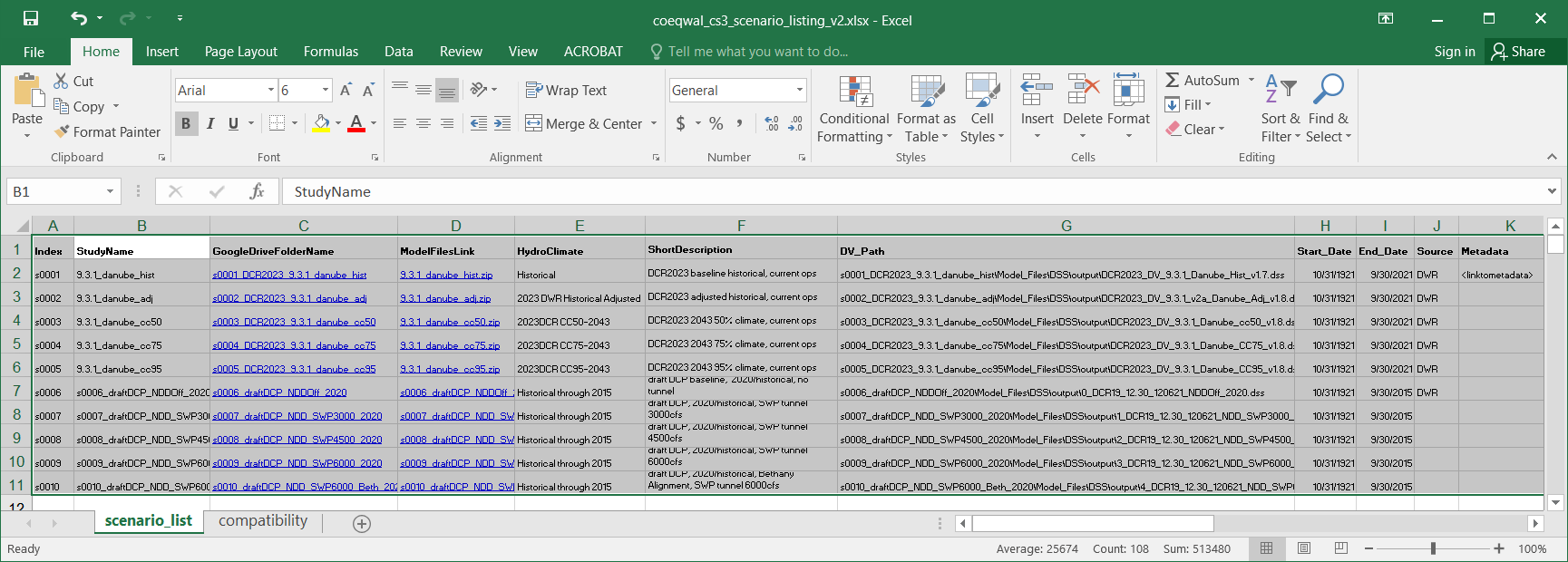
It needs to be downloaded to run all notebooks that extracts, manipulates, synthesizes, or plot CalSim3 data

* + - It contains a subdirectory **notebooks** that contains the current Jupyter notebooks, an initialization file (**CalSim3DataExtractionInitFile.xlsx**, discussed below), and a subdirectory ExcelDSSAdd-inPackage\_Feb2016 which contains libraries to interface DSS files.
      * **notebooks** in turn contains a subdirectory **coeqwalpackage** which contains python libraries referenced by the notebooks. It also contains a subdirectory

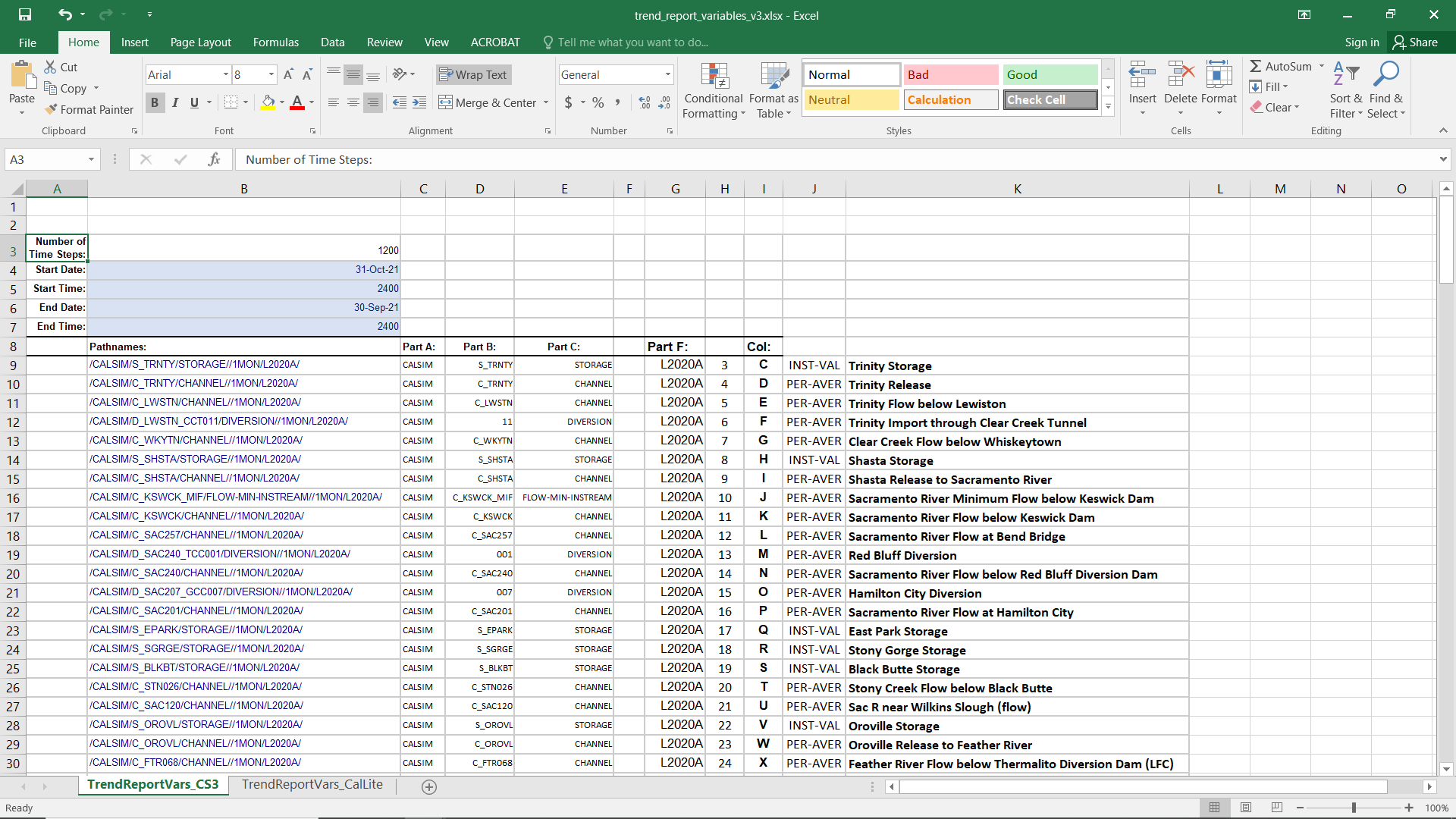
**Control files**

Three main files, mentioned above, are required to run the software, to identify the scenarios of interest, the variables of interest, and to initialize indices that point to files, tabs, variables, and attributes. In the current round of exploration, these are the following, but can be modified in the init file shown below:

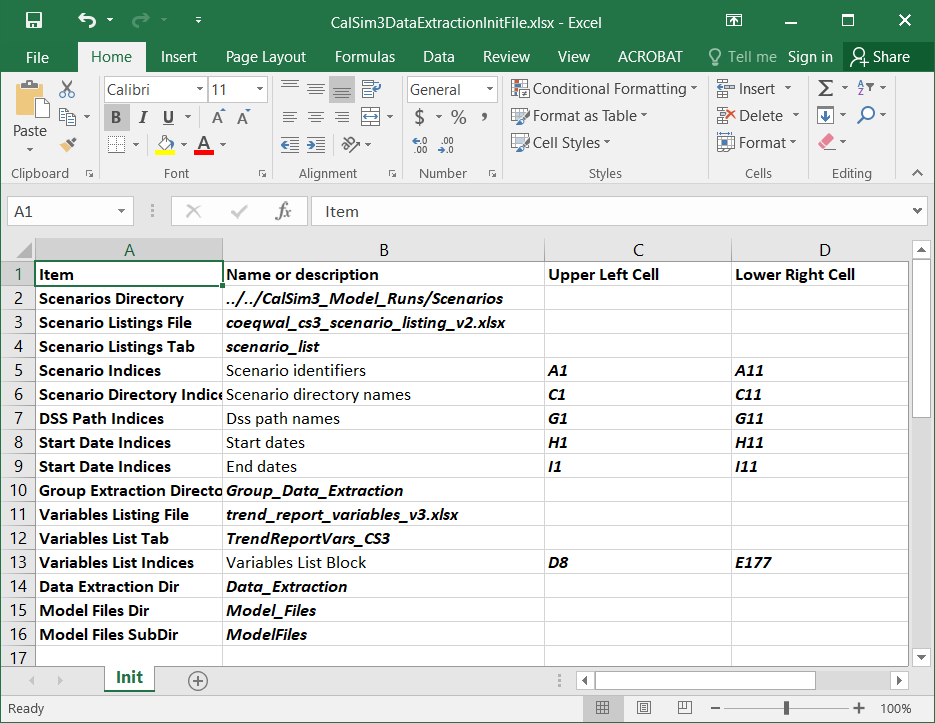
* **coeqwal\_cs3\_scenario\_listing\_v2.xlsx** which has the following format:



* **trend\_report\_variables\_v3.xlsx** which has the following format:



* **CalSim3DataExtractionInitFile.xlsx** which has the following format:

******

Note that the cells that are ***both bolded and italicized*** contain either file names, directory names, tab names, or cell block indices.

These represent the **current structure and current variables**. **These files needs to be modified when either the file structures or scenario or variable lists are modified!** The init file is then used to point to new files, tabs, scenarios, variables, etc.

**Initializing Jupyter notebooks**

Notebooks that use the above (or similar) structure only need to know the init file name and tab, and read the information contained in it. Notebooks also need to import the coeqwal package as well as the utilities to read from the excel files.

* To import modules:

*import sys*

*sys.path.append('./coeqwalpackage')*

*import cqwlutils as cu*

*from coeqwalpackage.DataExtraction import* \*

* To define file and tab names:

*CtrlFile = 'CalSim3DataExtractionInitFile.xlsx'*

*CtrlTab = 'Init'*

* To read info from the initialization file:

*DssListFile, DssListTab, DssListPath, DssNamesOutPath, DssIndicesOutPath, DssDirsOutPath, VarListPath, VarListFile, VarListTab, VarOutPath, DataOutPath, ConvertDataOutPath, ExtractionSubPath, ModelSubPath, GroupDataDirPath, ScenarioDir, DssMin, DssMax, DirMin, DirMax, IndexMin, IndexMax, StartMin, StartMax, EndMin, EndMax, VarMin, VarMax = cu.read\_init\_file(CtrlFile, CtrlTab)*

* For the above initialization file the variables will contain the following:

*print([DssListFile, DssListTab, DssListPath, DssNamesOutPath, DssIndicesOutPath, DssDirsOutPath, VarListPath, VarListFile, VarListTab, VarOutPath, DataOutPath, ConvertDataOutPath, ExtractionSubPath, ModelSubPath, GroupDataDirPath, ScenarioDir, DssMin, DssMax, DirMin, DirMax, IndexMin, IndexMax, StartMin, StartMax, EndMin, EndMax, VarMin, VarMax])*

['coeqwal\_cs3\_scenario\_listing\_v2.xlsx', 'scenario\_list', '../../CalSim3\_Model\_Runs/Scenarios\\coeqwal\_cs3\_scenario\_listing\_v2.xlsx', '../../CalSim3\_Model\_Runs/Scenarios\\Group\_Data\_Extraction\\DssNamesFrom\_coeqwal\_cs3\_scenario\_listing\_v2.csv', '../../CalSim3\_Model\_Runs/Scenarios\\Group\_Data\_Extraction\\IndicesFrom\_coeqwal\_cs3\_scenario\_listing\_v2.csv', '../../CalSim3\_Model\_Runs/Scenarios\\Group\_Data\_Extraction\\DirNamesFrom\_coeqwal\_cs3\_scenario\_listing\_v2.csv', '../../CalSim3\_Model\_Runs/Scenarios\\trend\_report\_variables\_v3.xlsx', 'trend\_report\_variables\_v3.xlsx', 'TrendReportVars\_CS3', '../../CalSim3\_Model\_Runs/Scenarios\\Group\_Data\_Extraction\\VarsFrom\_trend\_report\_variables\_v3.csv', '../../CalSim3\_Model\_Runs/Scenarios\\Group\_Data\_Extraction\\DataFrom\_trend\_report\_variables\_v3.csv', '../../CalSim3\_Model\_Runs/Scenarios\\Group\_Data\_Extraction\\ConvertDataFrom\_trend\_report\_variables\_v3.csv', 'Data\_Extraction\\Variables\_From\_trend\_report\_variables\_v3\_TrendReportVars\_CS3', 'Model\_Files\\DSS\\output', '../../CalSim3\_Model\_Runs/Scenarios\\Group\_Data\_Extraction', '../../CalSim3\_Model\_Runs/Scenarios', 'G1', 'G11', 'C1', 'C11', 'A1', 'A11', 'H1', 'H11', 'I1', 'I11', 'D8', 'E177']

* For example, to read the DSS pathnames:

*dsshdr, dssname = cu.read\_from\_excel(DssListPath, DssListTab, DssMin, DssMax, hdr=True)*

*dss\_names = []*

*for i in range(len(dssname)):*

*dss\_names.append(dssname[i][0])*

*dss\_names*

['s0001\_DCR2023\_9.3.1\_danube\_hist\\Model\_Files\\DSS\\output\\DCR2023\_DV\_9.3.1\_Danube\_Hist\_v1.7.dss',

's0002\_DCR2023\_9.3.1\_danube\_adj\\Model\_Files\\DSS\\output\\DCR2023\_DV\_9.3.1\_v2a\_Danube\_Adj\_v1.8.dss',

's0003\_DCR2023\_9.3.1\_danube\_cc50\\Model\_Files\\DSS\\output\\DCR2023\_DV\_9.3.1\_Danube\_cc50\_v1.8.dss',

's0004\_DCR2023\_9.3.1\_danube\_cc75\\Model\_Files\\DSS\\output\\DCR2023\_DV\_9.3.1\_Danube\_CC75\_v1.8.dss',

's0005\_DCR2023\_9.3.1\_danube\_cc95\\Model\_Files\\DSS\\output\\DCR2023\_DV\_9.3.1\_Danube\_CC95\_v1.8.dss',

's0006\_draftDCP\_NDDOff\_2020\\Model\_Files\\DSS\\output\\0\_DCR19\_12.30\_120621\_NDDOff\_2020.dss',

's0007\_draftDCP\_NDD\_SWP3000\_2020\\Model\_Files\\DSS\\output\\1\_DCR19\_12.30\_120621\_NDD\_SWP3000\_2020.dss',

's0008\_draftDCP\_NDD\_SWP4500\_2020\\Model\_Files\\DSS\\output\\2\_DCR19\_12.30\_120621\_NDD\_SWP4500\_2020.dss',

's0009\_draftDCP\_NDD\_SWP6000\_2020\\Model\_Files\\DSS\\output\\3\_DCR19\_12.30\_120621\_NDD\_SWP6000\_2020.dss',

's0010\_draftDCP\_NDD\_SWP6000\_Beth\_2020\\Model\_Files\\DSS\\output\\4\_DCR19\_12.30\_120621\_NDD\_SWP6000\_Beth\_2020.dss']

* For example, to read the scenario index names:

*indexhdr, index\_name = cu.read\_from\_excel(DssListPath, DssListTab, IndexMin, IndexMax, hdr=True)*

*index\_names = []*

*for i in range(len(index\_name)):*

*index\_names.append(index\_name[i][0])*

*index\_names*

['s0001',

's0002',

's0003',

's0004',

's0005',

's0006',

's0007',

's0008',

's0009',

's0010']

* For example, write a dataframe to the standard CSV file:

*df.to\_csv(DataOutPath, na\_rep="NaN", header=True, index=True)*

* For example, write a dataframe to the converted CSV file:

*df.to\_csv(ConvertDataOutPath, na\_rep="NaN", header=True, index=True)*

* For example, write a dataframe to the metrics\_output.csv in the group data extraction directory:

*out\_fp = os.path.join(GroupDataDirPath, “metrics\_output.csv)*

*df.to\_csv(out\_fp, na\_rep="NaN", header=True, index=True)*

**Jupyter notebooks**

There currently are three notebooks, which are used to extract data, convert units, and compute metrics. The latter also contains some plotting examples.

* They can be retrieved from the github repository [**https://github.com/maramahmedd/coeqwal**](https://github.com/maramahmedd/coeqwal). The extraction will create a local repository **coeqwal** in the same level as the local copy of the CalSim3 scenarios (e.g. **CalSim3\_Model\_Runs**)

**ExtractStudiesFromDssAndCompoundVariablesCalSim3.ipynb**

This notebook is fully controlled by the initialization file **coeqwal/notebooks/CalSim3DataExtractionInitFile.xlsx**

As explained above, this file determines where and what gets read/written. For simplicity, here we refer to filenames currently used in the init file, but note that these may be changed accordingly.

It extracts variables listed in **CalSim3\_Model\_Runs/Scenarios/trend\_report\_variables\_v3.xlsx** from the scenarios listed in the scenario list file **CalSim3\_Model\_Runs/Scenarios/coeqwal\_cs3\_scenario\_listing\_v2.xlsx**, and writes the variables from all listed scenarios tothe directory **CalSim3\_Model\_Runs/Scenarios/Group\_Data\_Extraction** in the file **DataFrom\_trend\_report\_variables\_v3.csv**. Note that this file name will change accordingly to the name of the variable listing file name. Important: the variable names have the scenario identifier (e.g. **\_s0001**) appended to specify their provenance.

It also extracts the variables from each individual scenario and places them in the directory **<scenario name>/Data\_Extraction** **Variables\_From\_<variable list file>\_<variable list tab>** (e.g. **CalSim3\_Model\_Runs\Scenarios\s0001\_DCR2023\_9.3.1\_danube\_hist\Data\_Extraction**\**Variables\_From\_trend\_report\_variables\_v3\_TrendReportVars\_CS3**). In this case the extracted variable file name reflects the name of the original DSS file that contains them (e.g. **DCR2023\_DV\_9.3.1\_Danube\_Hist\_v1.7.csv**).

**ConvertUnits.ipynb**

This notebook is also fully controlled by the initialization file **coeqwal/notebooks/CalSim3DataExtractionInitFile.xlsx**

It reads the extracted grouped variables from **DataFrom\_trend\_report\_variables\_v3.csv**, converts units from CFS to TAF if specified in **trend\_report\_variables\_v3.xlsx**, appends the converted columns to the data and writes it to the file **CalSim3\_Model\_Runs/Scenarios/Group\_Data\_Extraction**/**ConvertDataFrom\_trend\_report\_variables\_v3.csv**.

As in all the notebooks, file names may be changed in the initialization file

**Metrics.ipynb**

This notebook is fully controlled by the initialization file **coeqwal/notebooks/CalSim3DataExtractionInitFile.xlsx** with respect to data input. After reading the dataset, the user can subset the dataset based on variables and units, and compute different metrics, such as a suite of averages, and exceedance probabilities.

Plotting examples are also provided, including time series, month of year averages, and parallel plots (**Note: these are not yet included or documented!**)

Metrics and plots are written to subdirectories of **CalSim3\_Model\_Runs/Scenarios/Group\_Data\_Extraction** named **metrics\_output** and **plots\_output,** respectively.The file names have an appended time stamp, for easier identification.

Note that the selection of variables, metrics, and plots is left entirely to the user, but the notebook provides examples for guidance.