APEXMP Documentation

by Feng Pan © 2020 Updated: 20 January 2021

Available at: https://github.com/ryanpmcg/APEXMP

The **A**gricultural **P**olicy **Ex**tender Model **M**ultiprocessing **P**rogram (APEXMP) is an Python-based command line framework that was developed to supports automated data preparation, spatially-distributed (named Everycell) and semi-distributed (named Groupedcells) modeling as well as different results reporting. This framework is designed to conduct modeling for large spatial scales and/or fine spatial resolutions using parallel computing with the most current APEX model. While APEX version 1501 is the most recent model version, it is also expected that future updates to the core APEX model could be compatible with the proposed APEXMP framework. Although this was the original function of APEXMP to support APEX model only, it could be improved/modified that this framework will encourage greater utilization of other hydrologic models. This file provides a comprehensive documentation of the framework, its capabilities, the installation procedure, tutorials.

Contents

OVERVIEW

APEXMP Primary Purpose

APEXMP was designed with the main purpose of conducting Best Management Practices (BMPs) evaluations at field-scale with distributed architecture and automatic model setup of APEX. Parallelism drastically reduces computational time for spatially-distributed tasks. Included input database and automated functionality make APEX modeling efficient. The script is modular and can be readily maintained, modified, and expanded.

General Capabilities

Although the primary purpose for APEXMP is to, other capabilities were added to APEXMP such as . A comprehensive list of all methods is included in INTERNEL DESIGN section, but a broad summary of these capabilities is listed below:

Basic Capabilities (Always Performed)

- Included database and the data preparation module for the preparation of input data for setting up APEX.
- The main simulation module automatically updates input data to input files for the APEX model and finish setup and simulation.
- The gathering and mapping modules provides multiple options for analyzing, comparing, and displaying results.
- The automated visualization of the results at grid level by the map module illustrates the spatial distribution of NPS pollution and can identify areas with large contributions of pollution or areas with large potential of reducing pollution with BMP implementation.
- Parallel processing
- Export to File

Optional Capabilities (Performed with modification of input files)

- Replace the existing data with user own data.
- Adjust the parameters in APEX text files for optimization.
- Adjust the management files for comparing different BMPs.
- Adjust the APEX output control file for other results at different temporal scales or different hydrologic variables.

INSTALLATION

APEXMP Installation and Verification Procedure

There are a few steps that need to be completed before you are ready to run APEXMP. Make sure you have all the necessary files from the APEXMP GitHub: https://github.com/ryanpmcg/WEPPCLIFF. These files include: executable files for each components, required installation files, configuration files, documentation, a tutorial batch file (and inputs), APEX executable file, run APEXMP batch file, an installation verification test.

Once you have the files above, *clone the APEXMP master folder to the C:\ drive on your machine. If you are using another path, you must keep the path to this folder for later.* A tutorial follows, which assumes the folder was unzipped to the C:\ drive on a Windows machine. If you modify any step, you will be responsible for knowing how to proceed. Non-web-based installation alternatives are discussed at the end of this section.

Installation / Verification Outline

1. Install TauDEM and GDAL
2. Unzip the tutorial zip file
3. Run APEXMP tutorial and APEXMP batch

Installing TauDEM and GDAL

TauDEM 5.3.7 Complete Windows installer (This includes executables compiled for Windows, ArcGIS toobox and Python Scripts developed for ArcGIS 10.0 or higher, GDAL Library and Microsoft redistributables required to use MS-MPI.)
If you are using a Windows computer it is easiest to run the Complete Windows Installer (C:\APEXMP\INSTALL\ TauDEM537_setup.exe). During the install you will need to do the following:

- Click yes to the licensing agreements associated with the dependencies.
- When prompted to Choose Setup Type for GDAL, choose “Typical”.

Unzip the Tutorial zip file

Unzip the INPUTS.7z with any unzip software in your machine (preferrable 7-zip) with option “extract here” so it automatically put the “INPUTS” folder under “C:/APEXMP/”.

Using the Automated Tutorial for Verification

A tutorial batch file "AutoTutorial.bat" was provided to demonstrate the APEXMP program execution in four different scenarios which include both large and small study areas in distributed and semi-distributed configurations, respectively. To run the tutorial, simply double-click the "AutoTutorial.bat" batch file, and a command prompt window should appear while the jobs are being executed. Otherwise, you can choose to open the prompt window (cmd from windows option) as Administrator and then finish the two steps: paste “cd C:/APEXMP” and click enter; then paste “AutoTutorial.bat” and click enter. In the way, the prompt window will not be closed after the tutorial run finished.

Running the APEXMP batch with Customization of Configuration file

Change "C:/APEXMP/CONFIGS/DefaultConfig.txt" to your preferred options, place your "StudyArea.shp" shapefile in the "C:/APEXMP/INPUTS/StudyArea" folder, and then double-click the "RunAPEXMP.bat" batch file, and a command prompt window should appear while the job is being executed. Detailed procedures on how to customize a run are described in the TUTORIAL section.

The installation and verification time will vary based on your system hardware and internet connection. The process can take more than 10 minutes (for slow internet connections or older hardware) and as little as 30 seconds (for faster connections or newer hardware). Windows permissions sometimes prevents this step from working seamlessly, which is why the folder needs to be installed on the C:\ drive directly.

TUTORIAL

Basic Functions

Fig. 1 provides a visual representation of how APEXMP engages various model components based on user specified inputs. Users of the APEXMP framework are expected to know of the important differences of the Everycell and Groupedcells approaches and their respective advantages or limitations as no decision support is provided in the APEXMP framework to indicate which approach is more suitable for any particular simulation. Specific designs of each component are described in the INTERNEL DESIGN section.

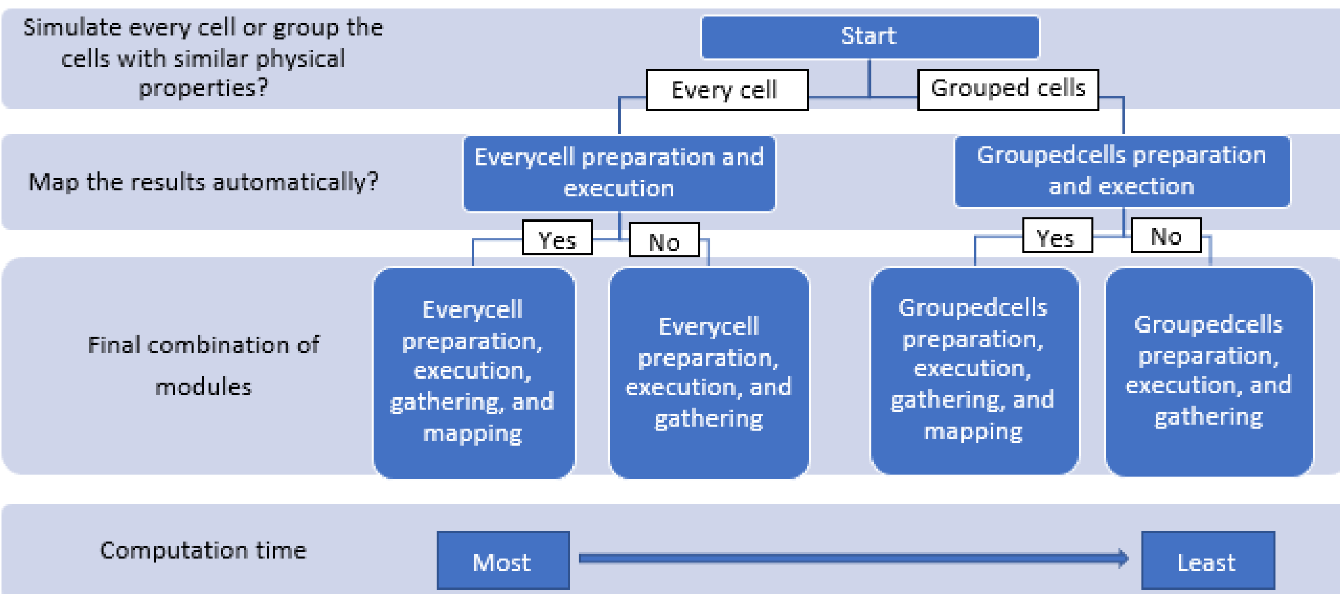


Figure 1. An overview of the APEXMP workflow, how various components are engaged based on user inputs, and how those options impact computation time.

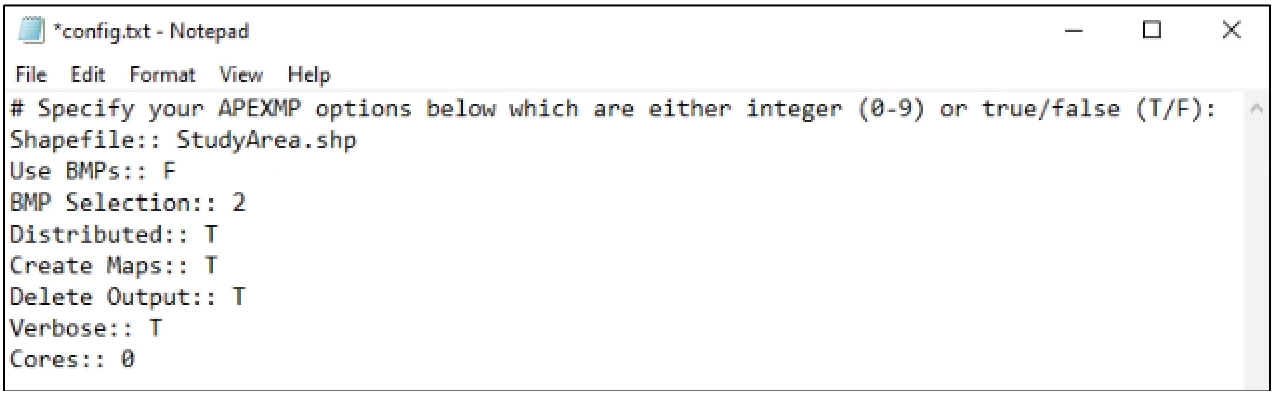
The following examples will introduce you to the basic functions of APEXMP. For example, this portion of the tutorial will show you how to customize the configuration file to conduct different basic build-in functions of APEXMP.

Double click the config.txt file to open it as shown in Fig. 2 and adjust the input variables:

1. For "Shapefile", change the "StudyArea.shp" to your study area shapefile name (already put in C:\APEXMP\INPUTS\studyarea) to conduct study in your desire area.
2. For "Use BMPs", if you want to compare the results among different BMPs, then replace "F" with "T" so that APEXMP will consider all the cells as one BMP at a time. Otherwise, the default setting "F" is to conduct simulation with all the actual crop management scenarios.
3. For "BMP Selection", if you input "T" for "USE BMPs", then input any number that you plan to compare plus one (one of them are baseline scenario), detailed methods of how to change the management scenario files are described in the INPUTS section. Otherwise, ignore this one.
4. For "Distributed", if you input "T", the everycell option will be chosen which costs more time

and potentially more accurate. If you input “F”, the groupedcells options will be chosen which costs less time and potentially less accurate. All based on your hardware and time to spend.

5. For “Create Maps”, “T” lets APEXMP create raster maps for each outputs and “F” does not.
6. For “Delete Output”, “T” lets APEXMP delete all intermedia data and APEX files or output to leave only necessary data, files, and results that can save significant storage. “F” makes APEXMP save all the data, files, and results it creates if you need them for some reason.
7. Ignore “Verbose”.
8. For “Cores”, leave it as “0” if you want to let APEXMP to decide the best number to use but not affect your other regular tasks on your PC. Otherwise, input any number of cores (must less or equal to your actual cores including virtual ones) that you think works best.


*config.txt - Notepad
File Edit Format View Help

```
# Specify your APEXMP options below which are either integer (0-9) or true/false (T/F):
Shapefile:: StudyArea.shp
Use BMPs:: F
BMP Selection:: 2
Distributed:: T
Create Maps:: T
Delete Output:: T
Verbose:: T
Cores:: 0
```

Figure 2. Screenshot of config.txt file.

Optional Functions

This section of the tutorial will show you how to conduct some optional functions of APEXMP.

1. Replace the existing data with user own data: all of the original raster data including DEM, soil, land use, slope, and zip code, and all of the climate data could be replaced if you want to use new/old version, finer/coarser spatial resolution, or other countries than US data. Detailed location and format of the data are described in the INPUTS section.
2. Adjust the parameters in APEX text files for optimization: adjust any parameters for calibration if you know the regular calibration procedures as for APEX simulation without APEXMP. Location of the files are in the INPUTS section.
3. Adjust the management files for comparing different BMPs: two management scenario files are provided as default setting that are fallow as baseline and pine tree as BMP; if you need more than this or need other BMPs and baseline, create your own scenario files with other APEX tools or manually following the format of APEX and also update the control file of scenario files. Location of the files are in the INPUTS section.
4. Adjust the APEX output control file for other results at different temporal scales or different hydrologic variables: only annual averages of runoff, soil loss, total N, and total P are included in the current version of APEXMP; you can adjust the output control file to have more temporal scales or other variables; the data preparation and model execution module will still be working fine but the current version of gathering and mapping modules won’t work as they are designed for specific results. Anyway, you will still have the raw results from APEX for analysis.

INTERNAL DESIGN

APEXMP includes four scripts ('modules') for each of the two major approaches (Everycell and Groupedcells). These include preparation, execution, gathering, and mapping modules. Fig. 3 shows the dataflow (inputs, processes, and outputs) of APEXMP. The whole framework was developed with the Python language supported by Pandas, Numpy, GDAL, Fiona and TauDEM packages.

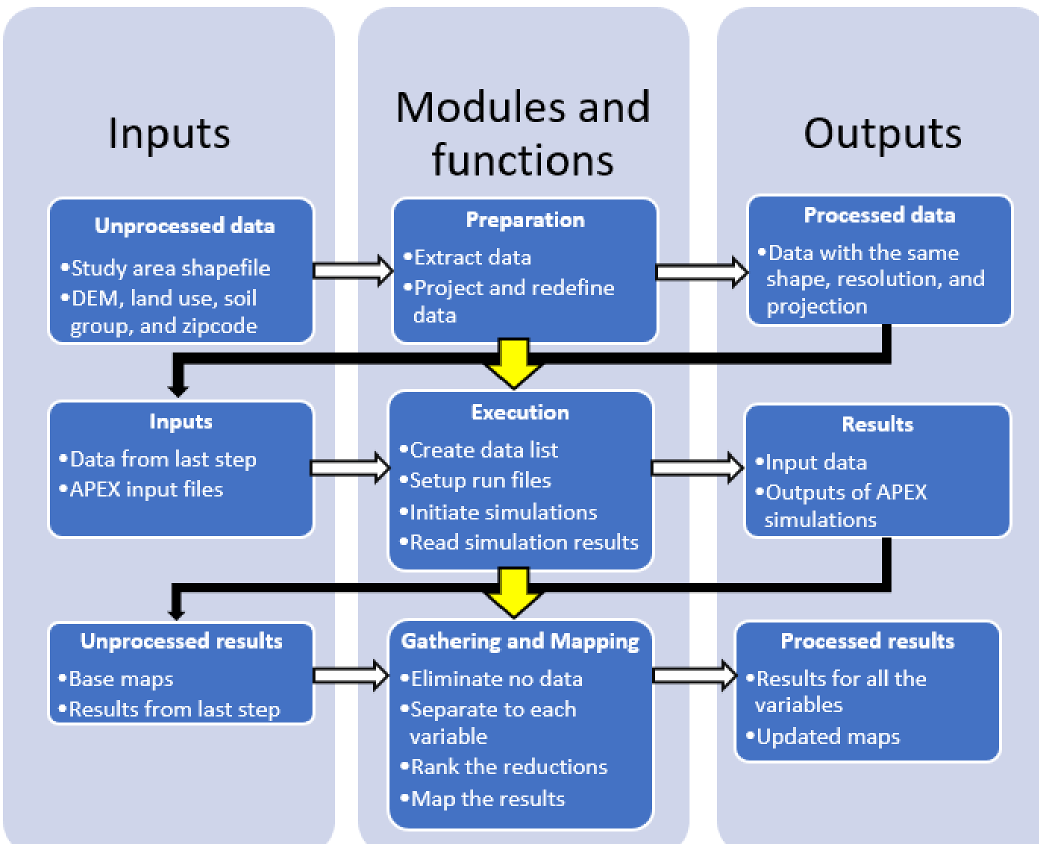


Figure 3. Dataflow in the APEXMP framework.

Data Preparation Modules

With everycell option, with the study area shapefile, the module uses the gdalwrap and gdal_translate functions of GDAL for clipping raster layers and converting the clipped raster layers from tiff files to ASCII datasets. The D8FlowDir of TauDEM library are applied to calculate slope using the DEM layer. With groupedcells option, the study area shapefile is used to clip the shapefile of counties for the study area with GDAL. The shapefile of counties in the study area are then disaggregated to shapefiles for each county with the Fiona library. The shapefiles for each county is then used as a mask to clip and transfer the raster layers as same as with everycell option with a looping function.

Model Execution Modules

The model execution modules consist of one central function and several sub-functions as shown in Fig. 4. The central function starts with calling the sub-functions of reading and overlapping multiple spatial layer information by creating a data list. The next part of the central function includes reorganizing the cellslist, creating an in-memory soil database, and creating a job pool for parallel model setup and runs. Model setup involves primarily the creation or retrieval of inputs for each job to be executed. The setup procedure calls multiple sub-functions to accomplish a range of tasks. The running procedure creates individual run logs and error files, initiates the model run for a given run folder, and finally removes unnecessary files to reduce the storage size. All setup functions are performed before model execution to avoid potential situations where a run setup encounters an issue but is not reported until after all simulations have been completed. The final part of the central function is reading and writing the input data and output results to a result file.

Gathering and Mapping Modules

As shown in Fig. 4, users can choose to process the results by either only gathering table outputs, or gathering table outputs and then mapping the results. With the no-map option, the result file(s) generated by the model execution modules are read, and the results with no data are eliminated. Next, the input data, simulation results, pollution reductions, and the rankings of the reductions are separated into multiple tabular files.

With the map and the groupedcells option, the results need to be disaggregated to individual cells with their own column and row numbers. The column and row numbers for each cell associated with different variables are first rearranged to Python Numpy arrays, and then update to the base map with GDAL open and write functions with the map option.

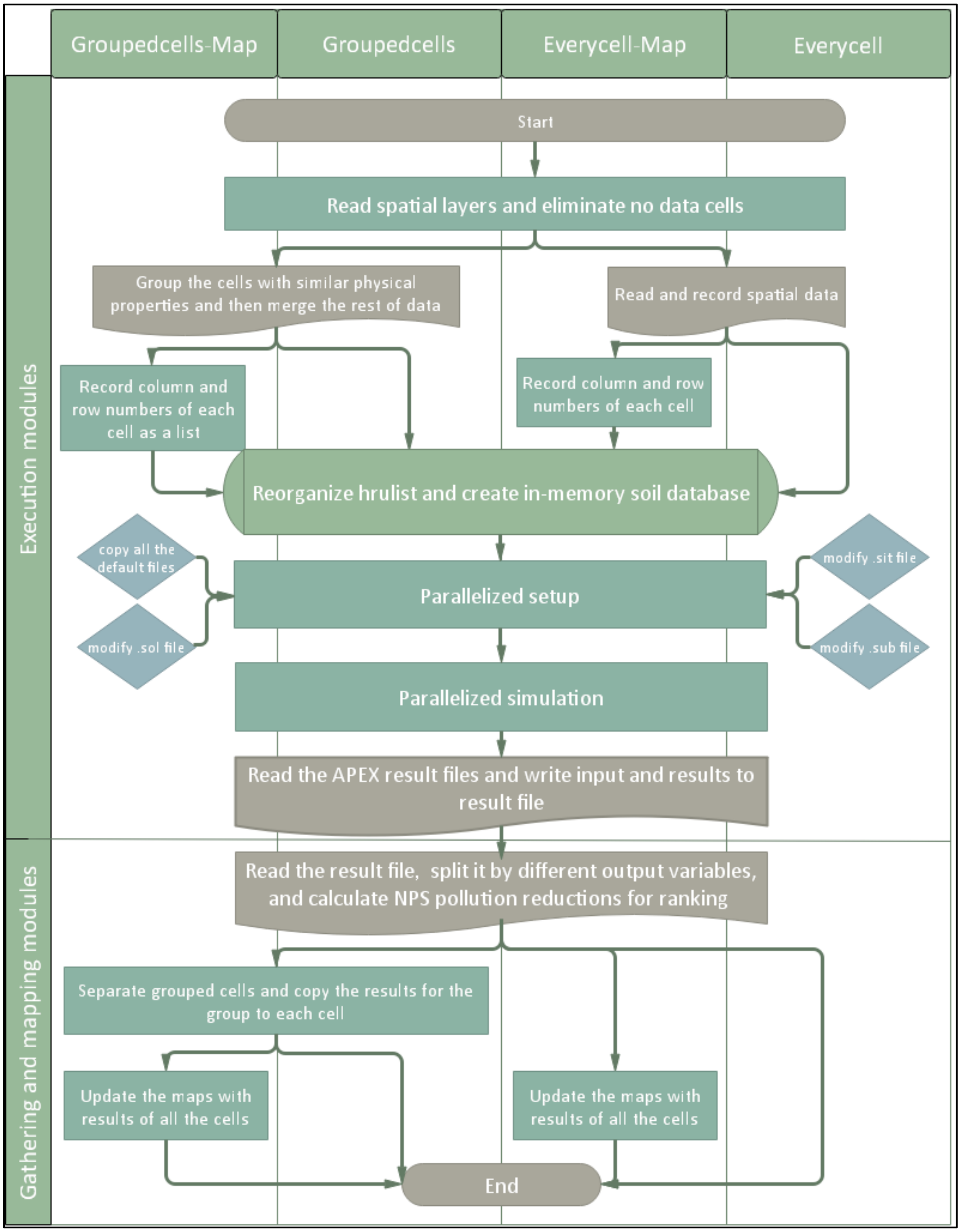


Figure 4. Conceptual workflows of the model execution, gathering and mapping modules for the four versions (blocks that cross all the versions represent shared processes; blocks that only exist in one or two versions represent special processes only for certain versions).

INPUTS

Input Data for Updates of APEX Input Files

Fig. 5 lists all of the original data or the data would be generated and their location after execution of the data preparation module. Users are required to provide a shapefile (YourShapefileNameHere.shp), which defines the study area and model boundaries. This is then used to extract information from the default APEXMP databases or (optionally) from a user's own database in the same format as the default database. At this time, default databases have been prepared for only the continental United States.

The Provided Database of the Continental US in the Package

1) Raster layers of elevation (alldem1.tif), land use (alllanduse.tif), soil (allsoil.tif), and zip codes (zipall.tif).

2) Vector layers of all US counties (stcty1.shp)

3) Map unit aggregate attribute table for soil group and county-zip code location (ssurgo2apex.csv)

4) State county-zip code-coordination tabular file (usstctyziplatlong2sql.txt)

All spatial layers have the same projection, and all the raster layers have the same resolution so that data can be extracted and overlapped spatially for updating APEX files. If a coarser or finer resolution is needed, users could prepare their own data and make them the same names, projection and resolution and put them in the same location as shown in Fig. 5.

The Generated Data for Updates of APEX Input Files

1) Counties shapefile for creating all shapefiles of counties (cty+YourShapefileNameHere.shp) and shapefiles of counties (in "./all")

2) Intermedia TIFF maps created from original maps and for ASCII maps (in "./indi")

3) Final ASCII maps (DEM, land use, soil, slope, and zip code) for the model execution module (in "./asc")

```
C:/APEXMP/INPUTS
          ./ssurgo2apex.csv__________________________(soil database, provided)
          ./usstctyziplatlong2sql.txt____________________(zip code database, provided)
          ./studyarea
                    ./YourShapefileNameHere.shp__________(study area shapefile, needed)
                    ./cty+YourShapefileNameHere.shp_______(study area county shapefile, generated)
          ./county
                    ./stcty1.shp________________________(includes all of the counties, provided)
                    ./all____________________________(includes all of the counties' shapefiles,
                                                       for Groupedcells version only, generated)
          ./dem
                    ./asc____________________________(final ASCII maps for the model execution module)
                    ./indi___________________________(intermediate tiff maps for converting to ASCII maps
                                                       , will be deleted at the end)
                    ./original
                              ./alldem1.tif________________(DEM for continental US, provided)
          ./landuse
                    ./asc
                    ./indi
                    ./original
                              ./alllanduse.tif ______________(cropland map for continental US, provided)
          ./soil
                    ./asc
                    ./indi
                    ./original
                              ./allsoil.tif__________________(soil group map for continental US, provided)
          ./zipcode
                    ./asc
                    ./indi
                    ./original
                              ./zipall.tif__________________(zip code map for continental US, provided)
          ./slope
                    ./asc
                    ./indi
                    ./original_________________________(slope map for continental US or
                                                       user specifies, optional)
```

Figure 5. Data and folder structure for original and processed input data (names without extensions are folders).

Default Input Files for APEX

Only those APEX input files that will be modified by APEXMP are described here as shown in Fig. 6. For more detailed information of each file, check the APEX manual also included in the package (THE-APEX1501-user-manual-March-2019.pdf) on how to modify these files.

1) APEX control file (APEXCONT.DAT) controls simulation time and period

2) APEX master file (APEXFILE.DAT) contains the locations of all the require files

3) APEX operation schedule list file (OPSCCOM_bmps.DAT for BMPs option and OPSCCOM_ops.DAT for no BMPs option) contains location and names of the operation files (OP*.OPC) that will be used in simulation

If users want to use their own land use data or BMPs management files, create or move you own APEX management files to here to replace the default management files, then make sure that the list numbers in the OPSCCOM.DAT must be corresponded to the land use number list or the BMPs list (i.e. if users choose "T" of "BMPs" and input "4" for "BMP selection", then four OP*.OPC must be listed as 1 to 4, or if choose "F" of "BMPs", number of land use types from land use map must be listed and corresponded number of OP*.OPC are needed here)

4) APEX soil, site, subarea list files (SITECOM.DAT, SOILCOM.DAT, and SUBACOM.DAT) contains the location and names of site, soil, and subarea files (SIT.SIT, SOL.SOL, and SUB.SUB) that will be used in simulation

These three files will be modified in the model execution module, so do not change them manually

5) APEX wind data list file (WINDWEPP.DAT) and monthly weather data list file (WPM1WEPP.DAT) contain list of wind and weather data stored in two folders of "./db03_wndus_weppcligen" and "./db02_wp1us_weppcligen"

Climate and wind data in these two folders can be replaced and then the corresponded numbers and names in the list files need to be changed accordingly

```
C:/APEXMP
        ./REQUIRED
                ./db02_wp1us_weppcligen____________(The monthly weather files)
                ./db03_wndus_weppcligen____________(The wind data files)
                ./APEXCONT.DAT___________________(APEX Control File)
                ./APEXDIM.DAT
                ./APEXFILE.DAT____________________(APEX Master File)
                ./APEXRUN.DAT
                ./CROPCOM.DAT
                ./CROPSWT.DAT
                ./FERTCOM.DAT
                ./FERTSWT.DAT
                ./HERD0806.DAT
                ./MLRN0806.DAT
                ./OPSCCOM_bmps.DAT______________(APEX Operation Schedule List File for BMPs)
                ./OPSCCOM_ops.DAT_______________(APEX Operation Schedule List File without BMPs)
                ./PARM0806.DAT
                ./PESTCOM.DAT
                ./PESTSWT.DAT
                ./PRNT0806.DAT
                ./PSOCOM.DAT
                ./RFDTLST.DAT
                ./SITECOM.DAT____________________(APEX Site List File)
                ./SOILCOM.DAT____________________(APEX Soil List File)
                ./SUBACOM.DAT___________________(APEX Subarea List File)
                ./TILLCOM.DAT
                ./TILLSWT.DAT
                ./TR55COM.DAT
                ./WDLSTCOM.DAT
                ./WINDWEPP.DAT__________________(APEX Wind Data List File)
                ./WPM1WEPP.DAT__________________(APEX Monthly Weather Data List File)
                ./OP*.OPC________________________(APEX Operation Schedule Files)
                ./SIT.SIT________________________(APEX Site File)
                ./SOL.SOL________________________(APEX Soil File)
                ./SUB.SUB________________________(APEX Subarea File)
```

Figure 6. Data and folder structure for APEX default input files (names without extensions are folders).

OUTPUTS

APEX Default Results

Depends on the user's selection of "distributed" in configuration file, the APEX default run folders store the results will be put in either "./MPRUNS/everyhru_map" for distributed option or "./MPRUNS/grouphru_map" for semi-distributed option as shown in Figure 7. Under folder "./MPRUNS/everyhru_map", the subfolders are named as run numbers starting from 0 and under folder "./MPRUNS/grouphru_map", the subfolders are named as state names included in the study area and then the next level subfolders are named as run numbers starting from 0. The results files will be read in the model execution module are "Results.WSS". Detailed output format can be found in "THE-APEX1501-user-manual-March-2019.pdf".

Model Execution Module Generated Results

Depends on the user's selection of "distributed" in configuration file, the gathered results by model execution module will be put in either "./RESULTS/everyhru_map" for distributed option or "./RESULTS/grouphru_map" for semi-distributed option as shown in Figure 7.

Table 1 shows part of an example result file (./RESULTS/everyhru_map/*.csv) from the model execution module with option "distributed" and includes run ID, six input data, operation number (if "BMPs" is "F", this will all be "1"), location ID (will be used for mapping), and five output variables.

Table 2 shows part of an example result file (./RESULTS/grouphru_map/*.csv) from the model execution module with option "semi-distributed" and includes run ID, state and county names, nine input data, location ID (will be used for mapping), operation number (if "BMPs" is "F", this will all be "1"), and five output variables.

Table 1. Example of APEXMP model execution module with option "distributed".

RunID	Soil	Landuse	Slope(%)	Latitude	Longitude	Elevation(m)	Operation	RowID_ColumnID	Precipitation(mm/yr)	Runoff(mm/yr)	Soil Loss(t/ha)	Total N(kg/ha)	Total P(kg/ha)
0	188666	1	3.557318	41.70007	-84.806	327.317	1	0_19	872.3	42.2	0.35	9.07	1.42
1	188666	1	3.557318	41.70007	-84.806	327.317	2	0_19	872.3	42.8	0.43	9.73	1.54

Table 2. Example of APEXMP model execution module with option "semi-distributed".

Run ID	State	County	Zip code	Soil	Land use	Slope group	Area (m2)	Slope (%)	Latitude	Longitude	Elevation (m)	Rowid_Colid	Operation	Precipitatio n (mm/yr)	Runoff (mm/yr)	Soil Loss (t/ha)	Total N (kg/ha)	Total P (kg/ha)
0	Indiana	Steuben	46737	160224	37	2	4500	0.384858	41.72157	-84.9219	329.053	2_7;3_7;3_8;4_14;4_15	1	863.6	35	0.18	1.05	0.75
1	Indiana	Steuben	46737	160224	37	2	4500	0.384858	41.72157	-84.9219	329.053	2_7;3_7;3_8;4_14;4_15	2	863.6	0.7	0	0.02	0.01

Gathering and Mapping Modules Generated Results

Depends on the user's selection of "distributed" in configuration file, the processed tabular results by gathering module will be put in either "./results analysis/everyhru" for distributed option or "./RESULTS/grouphru" for semi-distributed option as shown in Figure 7.

Table 3 shows part of an example result file (./results analysis/everyhru/*.csv) from the gathering module with option "distributed" and includes run ID, coordinates, location ID (will be used for mapping), specific output variable for each BMPs (if "BMPs" is "F", this will only one output), reduction from BMPs to baseline (Scenario 1 is baseline), and the ranking among all of the differences.

Table 4 shows part of an example result file (./results analysis/grouphru/*.csv) from the gathering with option “semi-distributed” and includes run ID, location ID (will be used for mapping), specific output variable for each BMPs (if “BMPs” is “F”, this will only one output), reduction from BMPs to baseline (Scenario 1 is baseline), and the ranking among all of the differences.

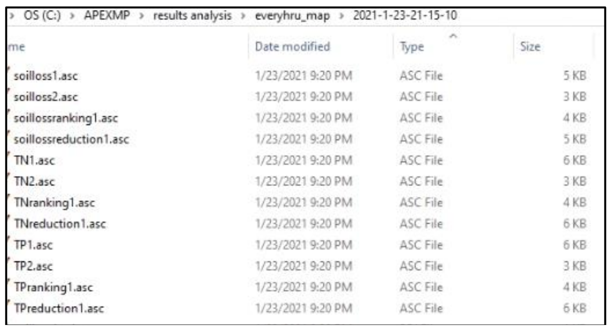
Table 3. Example of APEXMP gathering module with option “distributed”.

RunID	RowID_ColumnID	Longitude	Latitude	Soil Loss1(t/ha)	Soil Loss2(t/ha)	op2-op1	op2-op1_ranking
0	0_18	-84.9335	41.47035	1.88	0	-1.88	179
2	0_19	-84.9332	41.47035	1.71	0	-1.71	201

Table 4. Example of APEXMP gathering module with option “semi-distributed”.

RunID	Soil Loss1(kg/ha)	Soil Loss2(kg/ha)	Rowid_Colid	op2-op1	op2-op1_ranking
0	2.39	0	0_18	-2.39	71
2	2.39	0	0_19	-2.39	72

Depends on the user’s selection of “distributed” in configuration file, the processed tabular results by mapping module will be put in either “./results analysis/everyhru_map” for distributed option or “./RESULTS/grouphru_map” for semi-distributed option as shown in Figure 7. Maps will be created for all of the combination of variables, scenarios, reductions, and rankings in both TIFF and ASCII formats (*.tif and *asc) as shown in the example screenshot below.




```
C:/APEXMP
    ./MPRUNS
        ./everyhru_map____________________(run folders with everycell option, generated)
        ./grouphru_map____________________(run folders with groupedcells option, generated)

    ./RESULTS
        ./everyhru_map
            ./*.csv____________________(all files of results with everycell option)
        ./grouphru_map
            ./*.csv____________________(all files of results with groupedcells option)
    ./results analysis
        ./everyhru_map
            ./*_______________________(results directory with csv file name
                                       with everycell option, generated)
                ./*.tif_______________(TIFF maps with everycell option, generated)
                ./*.asc_______________(ASCII maps with everycell option, generated)
        ./everyhru
            ./*_______________________(results directory with csv file name
                                       with everycell option, generated)
                ./*.csv_______________(spreadsheet results files with
                                       everycell option, generated)
        ./grouphru_map
            ./*_______________________(results directory with csv file name
                                       with groupedcells option, generated)
                ./*.tif_______________(TIFF maps with groupedcells option, generated)
                ./*.asc_______________(ASCII maps with groupedcells option, generated)
        ./grouphru
            ./*_______________________(results directory with csv file name
                                       with groupedcells option, generated)
                ./*.csv_______________(spreadsheet results files with
                                       groupedcells option, generated)
```

Figure 7. Data and folder structure for results and results analysis (names without extensions are folders).

CONTRIBUTORS

Acknowledgements

As of the most recent release of APEXMP, Component scripts were written by Feng Pan, Ryan McGehee, and Qingyu Feng (in order of contribution).

Within APEXMP there are several packages used from the Python community. All of these important packages are listed here:

- GDAL
- TauDEM
- Fiona
- Pandas
- Numpy

Citation for Published Works

Pan, F., Q. Feng, R.P. McGehee, B. Engel, D.C. Flanagan, J. Chen. 2021. A parallel modeling framework for spatially-distributed modeling with the Agricultural Policy Environmental eXtender (APEX) model. Environmental Modelling and Software (Submitted).

Contact Information

Feng Pan, Lead Developer Email:
pan315@purdue.edu