DISTRIBUTED HYDROLOGICAL MODEL FOR WATERSHED MANAGEMENT

(DHM-WM)

USERS' MANUAL

VERSION 2018

SISI LI

April, 2018

INSTITUTE OF GEODESY AND GEOPHYSICS,

CHINESE ACEDEMY OF SCIENCES

Software Preparation

1. Python 2

The scripts for DHM-WM is written in Python programming language. Since some of the needed packages are based on Python 2, the model script is also written in Python 2. The Python version currently used is 2.7.14.

2. Python packages

DHM-WM uses some open-source python packages to do data analysis, including: NumPy, Pandas, GDAL, PyGeoprocessing. The detail introduction of these packages and how to install them can be found by searching them in the PyPI (https://pypi.org/) or the homepages of these packages. The usage and currently-used versions in DHM-WM are described as below.

(1) NumPy

NumPy is the fundamental package for scientific computing with Python (http://www.numpy.org). DHM-WM mainly uses it to do raster calculation. The currently used version is 1.13.3. Since NumPy is the dependency of Pandas, their versions should be compatible. It is recommended to install NumPy and Pandas together, such as installing via pip, typing in the commands:

```
python -m pip install --user numpy pandas
```

Note: If you have both python 2 and python 3 installed, use py -2 instead of python in theabove command, to specify the python version that you would like to install the NumPy package.

(2) Pandas

Pandas is a library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language (http://pandas.pydata.org). DHM-WM mainly uses it to deal with time series, such as climate data. The currently used version is 0.21.0.

(3) GDAL

GDAL is a translator library for raster and vector geospatial data formats (http://www.gdal.org). DHM-WM mainly uses it to read and write spatial data. The currently used version is 2.2.2. Its installation on Windows is a bit complex, it will be helpful to read the instructions here (http://cartometric.com/blog/2011/10/17/install-gdal-on-windows/).

(4) PyGeoprocessing

PyGeoprocessing is a Python/Cython based library that provides a set of commonly used raster, vector, and hydrological operations for GIS processing. DHM-WM uses it mainly for hydrological operations instead of ArcGIS functions. The routing libraries of versions from 0.4.1 to 0.6.0 have some bugs, so the version 0.3.3 is currently used in DHM-WM. It can be easily installed via pip, but you need to install Shapely first, which Pygeoprocessing is based on.

You can check whether the above packages are successfully installed by import them in the Python Shell as below:

Python 2.7.14 Shell

File Edit Shell Debug Options Window Help

```
Python 2.7.14 (v2.7.14:84471935ed, Sep 16 2017, 20:25:58) [MSC v.1500 64 bit (AM
D64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>> import numpy as np
>>> import pandas as pd
>>> np.__version__
'1.13.3'
>>> pd.__version__
u'0.21.0'
>>> from osgeo import gdal
>>> from osgeo import ogr
>>> from osgeo import osr
>>> from pygeoprocessing import geoprocessing
>>> from pygeoprocessing import routing
>>>
```

1. Watershed Preparation

Watershed preparation is a preprocessing procedure before DHM-WM runs. In this procedure, watershed is delineated with DEM, outlet point (and watershed boundary) data; and flow direction raster is created, making sure water flows into streams. A script named “Watershed_preparation.py” is provided to do the preprocessing.

1.1 Input Data

The input data used to do watershed preparation include: a GeoTIFF format raster representing the DEM that covers the study watershed (the elevation unit should be meters), a point ArcGIS Shapefile representing watershed outlet point, a polyline Shapefile representing watercourses (and a polyline Shapefile representing watershed boundary if available). The boundary data is optional; it is used to set the boundary of the delineated study watershed, but can cover a larger area. The outlet point should be within the boundary.

Users should set the path of a folder that these input data are stored (in a variable **inputFolder**, line 21). Raster data, i.e., “dem.tif” should be stored in a subfolder (named “rasters”) within the **inputFolder**; vector data, i.e., “boundary.shp”, “outlet.shp”, “streams.shp” should be stored in a subfolder (named “vectors”) within the **inputFolder**.

Notes:

(1) Whether boundary data is available can be set (in the variable **boundaryData**, line 36), the default is “True”. If boundary data is not available, set it as “False”.

(2) If you use other names for the input data, change relative names in scripts (variables **dem_uri**, **str_uri**, **outlet_uri**, **boundary_uri**, line 32-37)

1.2 Input Parameter

One parameter (variable **str_crt**, line 40) is used to define whether a grid cell is regarded as a stream. If the number of the upstream grid cells is larger than the value of **str_crt**, the grid cell is defined as a stream or a watercourse. The smaller the value of **str_crt**, the denser the stream/watercourse networks. You can change the value of **str_crt** to adjust delineated stream/watercourse networks, making it more representative to the real one (your input data, such as “streams.shp”).

1.3 Output Data

You can run the “watershed preparation.py” script now by clicking “run”—“run module”, see Figure 1-1. When it is completed, an indication text will prompt out:

“*Watershed preparation is completed.*

Please check watershed and stream features created.

You can adjust str_crt to recreate stream feature, making it more representative to the real flowlines of the watershed."

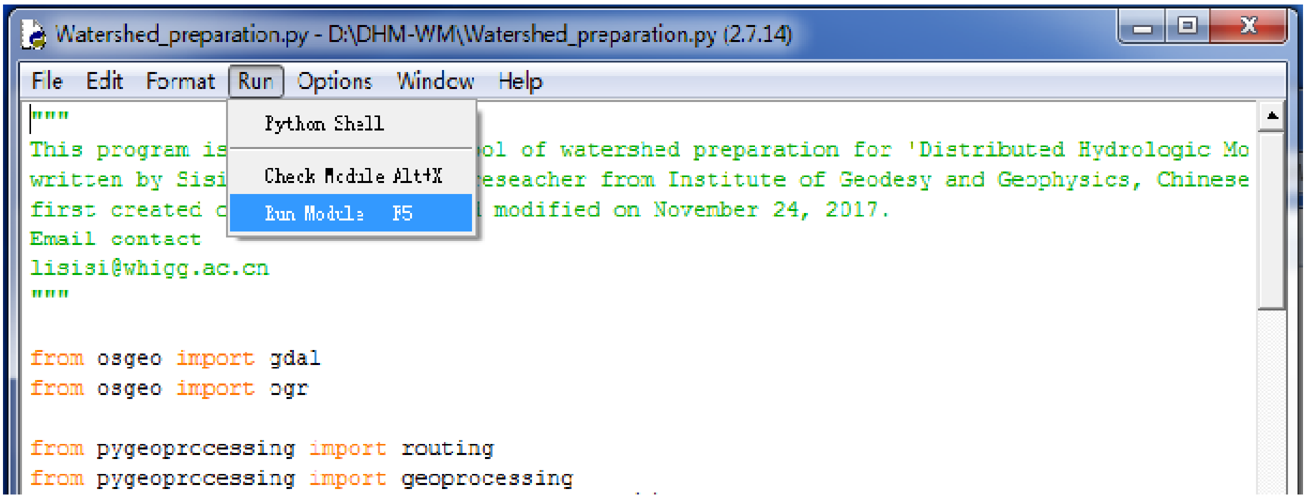


Figure 1-1. Run “Watershed preparation”

The output data include: “watershed.shp”, the delineated study watershed from input data; “stream_DL.shp”, the delineated stream/watercourse networks that guarantees water flowing into watercourses; and “fdr.tif”, the flow direction raster from watershed analysis. These output data are also stored in the inputFolder: rasters in “rasters” subfolder, vectors in “vecotrs” subfolder, see Figure 1-2.

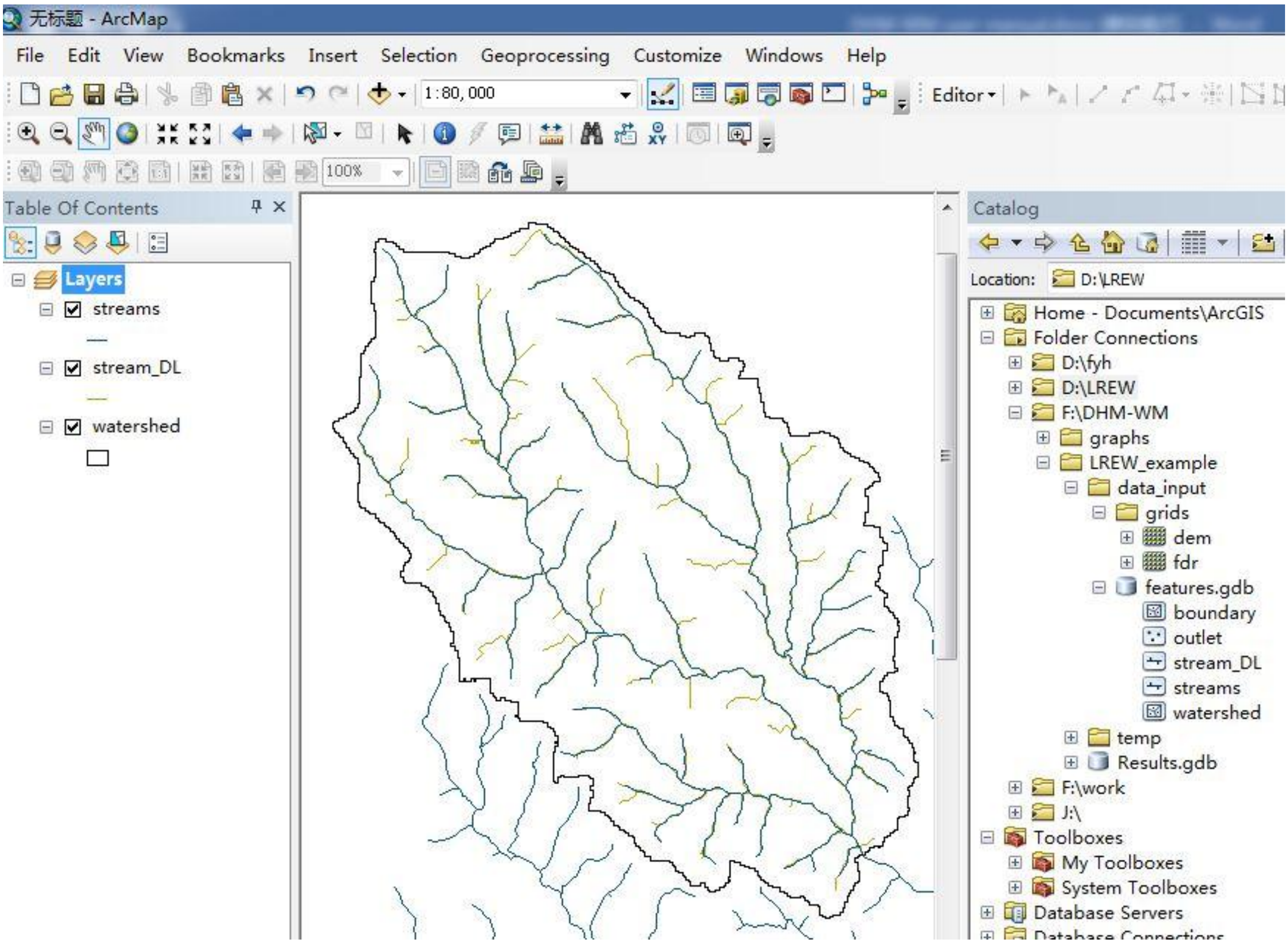


Figure 1-3. Example of output data

You can check whether “watershed.shp” and “stream_DL.shp” is proper. If the density of watercourses in “stream_DL” is quite different from that in “streams.shp” input data, adjust parameter **str_crt** and redo the analysis. When the density of the input “streams” data and the output “stream_DL” data is similar, as shown in the example in Figure 1-2, keep the output data.

However, there may still be some difference between the input “streams.shp” data and the output “stream_DL.shp” data, you can modify the output data “stream_DL” by editing in ArcGIS. Note: Do not manually add polylines into “stream_DL.shp” or move the location of polylines, just delete some extra ones. Meanwhile, you can merge several polylines into one if they represent the same stream reach. Last, you should add a field named “strWidth” to the property table of “stream_DL.shp”, and input stream width (in meters) data for each stream reach for further analysis. The modified “stream_DL” is as Figure 1-4.

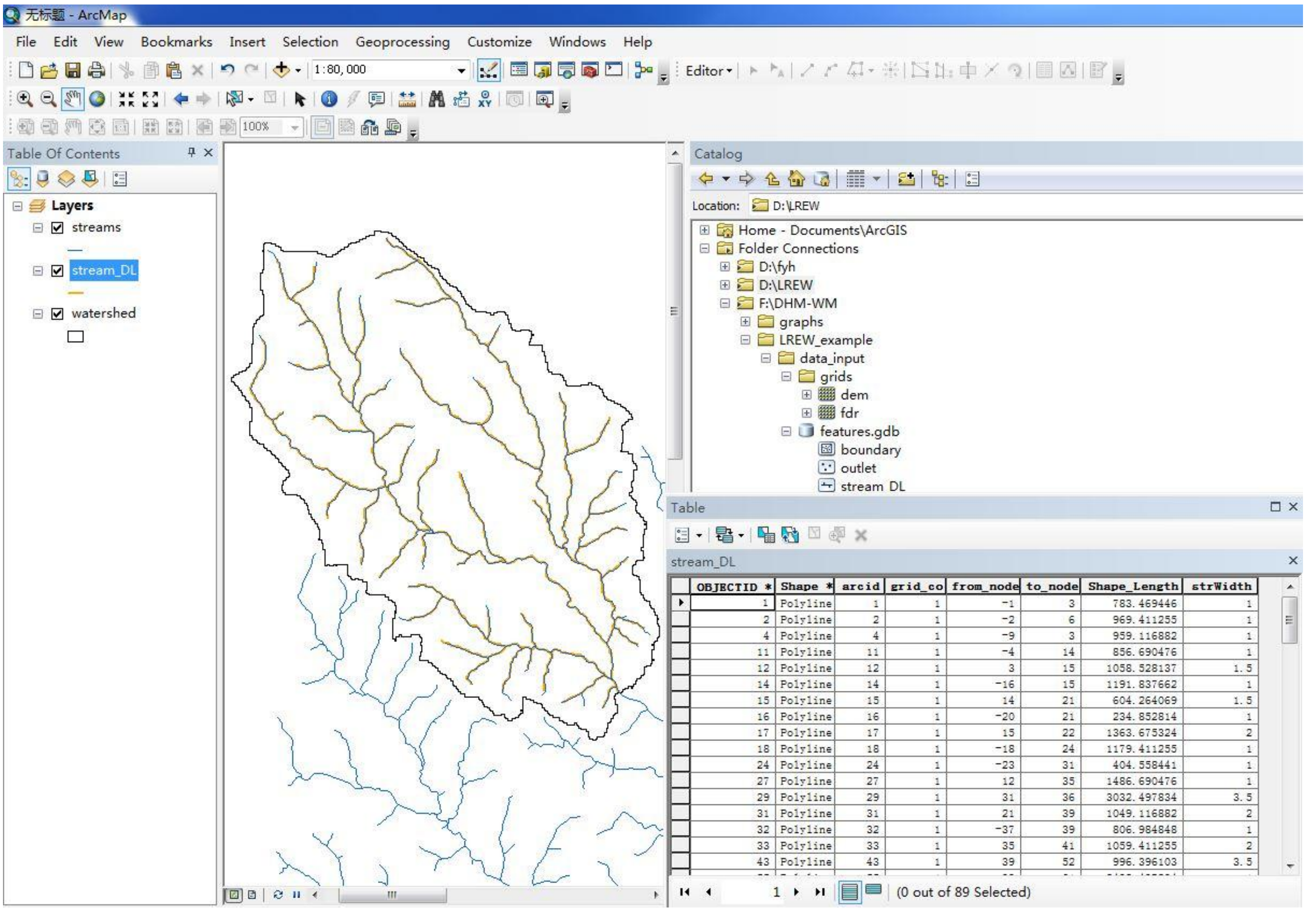


Figure 1-4. Modified “stream_DL.shp” that matches “streams.shp” better and its property table

2. DHM-WM Local Water Balance Routine

2.1 Input Data

The input data for DHM-WM local water balance routine include spatial data and text files, which are stored separately in two folders. Users should set the path of a folder that spatial data are stored (in a variable **inputFolder**, line 26); and the path of a folder that text files are stored (in a variable **python_inputFd**, line 28).

2.1.1 Input spatial data

The spatial data that users need to input include: raster data in GeoTiff format (flow direction, soil, land use, DEM, and gridded precipitation data if available); and vector data in ArcGIS Shapefile format (delineated watershed, and delineated streams from watershed preparation analysis, output sites for discharge simulation). Raster data should be stored in a subfolder named “rasters” within the **inputFolder**; and vector data should be stored in a subfolder named “vectors” within the **inputFolder**. These data are described in detail below.

(1) Flow direction data named “fdr.tif”, which is the output data of watershed preparation and should be stored just in the “rasters” subfolder. (Variable **fdr_uri**, line 47)

(2) Soil data named “soil.tif”, which is the spatial distribution of the soil type. The “value” field of the raster will be used to connect to their properties data which are prepared in a text file. “soil.tif” should cover the extent of the delineated study watershed. (Variable **soil_uri**, line 48)

(3) Land use data named “landuse.tif”, which is the land use/land cover raster of the study watershed. The “value” field of the raster will be used to connect to their properties or parameters which are prepared in a text file. The “landuse.tif” data should cover the extent of the delineated study watershed. (Variable **landuse_uri**, line 49)

(4) DEM data named “dem.tif”, which is the same as the input data of “watershed preparation” analysis. The elevation unit should be meters. (Variable **dem_uri**, line 50)

(5) Delineated watershed vector from watershed preparation analysis named “watershed.shp”, which is the output data from the watershed preparation. (Variable **watershed_uri**, line 51)

(6) Delineated stream vector from watershed preparation analysis named “stream_DL.shp”, which contains a field named “strWidth” that records the stream width data in meters. (Variable **stream_uri**, line 52)

(7) Output sites for discharge simulation named “discharge_sites.shp”, which are the target sites that need to report discharge time series. Generally, the more the number of target sites, the longer it will take for model computation. The points in “discharge_sites.shp” data should lie on the watercourse polylines of “stream_DL.shp” data. Each point in the “discharge_sites.shp” should have a field named “name” which describe the name of each site, the simulating results will be reported using these names. (Variable **sitesDischarge_uri**, line 53)

    In addition, the watershed outlet should be one of the points in “discharge_sites.shp”. The name of the watershed outlet is provided in Variable **outlet_name** in line 54 (as “AXL” in the example)

(8) Gridded precipitation data, which are optional. If gridded precipitation data are available, set the variable **haveDisPrcp** (line 56) to be True, and set the name of the folder where gridded precipitation data stores (Variable **DisPrcpFolder**, line 58). The folder should be within the sub-folder "rasters" of **inputFolder**, and the default name is "prcp/". The gridded precipitation data should be named as "prcpYYYYMMDD.tif" such as "prcp20030101.tif". In the AXL_example, gridded precipitation data is not available, but in the LREW_example, gridded precipitation data are used. In LREW_example, the gridded precipitation data can be interpolated from monitoring data at precipitation gauges by Inverse Distance Weighting (IDW) method. In the folder "LREW_example\data_input\grids\prcp\point data", there is a script "precip_point2grid.py" that can do the interpolation. The gauge data are prepared in ArcGIS .xyz format and stored in the same folder.

2.1.2 Input text files

The text files that need for DHM-WM local water balance routine are stored in the folder defined by variable **python_InputFd** (line 28), and include:

(1) Land use related properties and parameters named "landuse.txt", which include columns named "Value", "Name", "Manning_n", "CN_A", "CN_B", "CN_C", "CN_D". "Value" column is used to connect to the "Value" field of "landuse" raster data. "Name" column is a 4-character upper case symbol to imply the land use type, common "Name"s are listed in Appendix table 1. "CN_A", "CN_B", "CN_C", "CN_D" are the default curve numbers for

Hydrologic Soil Group (HSG) type A, B, C, D, respectively. The default curve numbers are determined based on NRCS National Engineering Handbook (NEH) 630 table 9, see the file in the folder of "related materials". (Variable **landuseFile**, line 62)

(2) Soil type related properties named "soil.txt", which include columns named "Value", "HSG", "Soil_z(m)", "Soil_k(m/day)", "Pe", "Pt". "Value" column is used to connect to the "Value" field of "soil" raster data. "HSG" is the HSG type defined by NRCS, which can be directly extracted from SSURGO database. Note: the values of "HSG" columns should be numbers, "HSG" value 1 refers to type A, value 2 refers to type B, value 3 refers to type C, value 4 refers to type D. "Soil_z(m)" is the soil depth to impervious/restrictive layer (in the unit of meters), which is determined by checking the saturated hydraulic conductivity with depth. Impervious/restrictive layer is defined as one with conductivity below 10 mm h^{-1}. "Soil_k(m/day)" is the average saturated hydraulic conductivity of the pervious soil profile (in the unit of meters per day), which is the weighted average saturated hydraulic conductivity of different layers above impervious/restrictive layer. "Pe" and "Pt" are the average effective porosity and total porosity of the pervious soil profile, which can be determined with other soil properties, such as sand-silt-clay percentages. Two methods are recommended to estimate "Pe" and "Pt" (see "Porosity estimation" in "related materials"): (a) a simple porosity calculator based on a reference file from Argonne laboratory, which need only sand-silt-clay percentages; (b) a soil water characteristics tool of SPAW hydrology developed by USDA ARS incorporated with Washington State University (https://hrsl.ba.ars.usda.gov/SPAW/Index.htm), which needs sand-silt-clay percentages, organic matter content, salinity and compaction. (Variable **soilFile**, line 63)

(3) Daily precipitation data named “prcp_d.txt”, which contains two columns: “date” and “prcp_d(mm)”. Column “date” is in the format of “YYYY-M(M)-D(D)” and should cover all simulation periods, including warm-up period, calibration period and validation period. Do not contain missing values, fill it as reasonable as possible. “prcp_d(mm)” is the daily precipitation in mm. (Variable **prcp_File**, line 65)

(4) 12-h and 24-h synthetic rainfall distribution data named “12raindis.txt” and “24raindis.txt”, both of which contain two columns: “hour” and “cmmlPrcp”. These data are used to estimate rainfall distribution within a day and calculate the duration of effective rainfall or runoff generation: use “12raindis.txt” when daily precipitation is below 50 mm and “24raindis.txt” when daily precipitation is above 50 mm. “cmmlPrcp” represents the cumulative precipitation rate at a specific time “hour”. Synthetic rainfall distribution can be estimated with rainfall frequency data using a spreadsheet tool provided by USDA NRCS (“Rainfall_Distribution_Spreadsheet_Tool.xlsx” in the folder of “related materials”). The sheet “Readme” in this tool provides links to information on rainfall, frequency and distribution, and where frequency data can be downloaded. “12raindis.txt” and “24raindis.txt” can be prepared with 3 steps. Step 1, calculate synthetic rainfall distribution with frequency data. Input rainfall frequency data into the “rainfall – in” column (C8-C17) in Sheet 1. Step 2, transform duration from 24-h to 12-h. Copy the estimated “Ratio” column (B45-B285) in Sheet 1 to “12-hour transformation” (column K) in sheet “duration transformation”, note: Select the proper rainfall distribution type by clicking specific type in row 2, and paste the “Ratio” in proper place. Step 3, transform results from 0.1h increment to 0.5h increment. Copy the estimated “12-hour distr.” (column N, in pink color) to the

orange area in sheet “increment transformation”, and the columns D-E are the 12-hour rainfall distribution that you can copy to the “12raindis.txt” file; copy the 24-h “Ratio” column (B45-B285) in Sheet 1 that you have copied in step 2 again and paste to the red area (Column I) in sheet “increment transformation”, and the columns K-L are the 24-hour rainfall distribution that you can copy to the “24raindis.txt” file. (Variable **rainDis12_File** in line 66, and **rainDis24_File** in line 67)

(5) Daily temperature data named “tmp_d.txt”, which contains columns named “date”, “TEM_max”, “TEM_min”. Column “date” is in the format of “YYYY-M(M)-D(D)” and should cover all simulation periods, including warm-up period, calibration period and validation period. Columns “TEM_max”, “TEM_min” represent maximum air temperature, minimum air temperature, respectively, in degrees Celsius. Missing values are represented as -99.9. Several missing values are OK, but more than 7 consecutive missing values may result in bias. (Variable **tmp_File**, line 68)

(6) Daily potential evapotranspiration (PET) data named “PET_d.txt”, which is optional. If observed PET data is available, set variable **haveDailyPET** to be “True” (line 69), and prepare “PET_d.txt” (Variable **PET_File**, line 71) in folder **python_inputFd**. There should be two columns in “PET_d.txt” file: “date” and “PET(mm)”. Column “date” is in the format of “YYYY-M(M)-D(D)” and should cover all simulation periods, including warm-up period, calibration period and validation period. Column “PET(mm)” is the PET of specific day in the unit of mm. Missing values are allowed in “PET_d.txt” file and the value of “PET(mm)” should be set to “-6999”. If observed PET data is unavailable, users need to set variable

haveDailyPET to be "False", and set the method used to calculate daily PET (variable **PET_method** in line 73). DHM-WM provided two methods: 1 represents Hargreaves method and 2 represents FAO Penman-Monteith method.

(7) Daily wind speed, sunlight duration, and humidity data, if **PET_method** is set to be 2 (FAO Penman Monteith). These data are needed for PET calculation using FAO Penman Monteith method. The wind speed data file named "wind_d.txt" (variable **win_File** in line 75) should contain at least two columns: "date" and "WIN_ave" (daily average wind speed measured at 10 m height, in m/s), missing values are represented as -99.9; the sunlight duration data file named "SSD_d.txt" (variable **SSD_File** in line 76) should contain two columns: "date" and "SSD" (daily actual sunlight duration, in hr), missing values are represented as -99.9; the humidity data file named "RHU_d.txt" (variable **humidity_File** in line 77) should contain at least two columns, "date" and "RHU_ave" (daily average relative humidity, in %), missing values are represented as -99.

(8) Daily crop coefficients (Kc) of a year for different land covers named "Kc_d.txt", which is optional. If the user would like to account for the effect of vegetation on PET, daily crop coefficients (Kc) would be used to adjust PET for different land cover and different day. In this case, set variable **haveDailyKc** (line 79) to be True and prepare the "Kc_d.txt" data (variable **Kc_d_File**, line 81). The "Kc_d.txt" file has 366 rows representing the day of year. The number of columns is the number of land use types of the study watershed, and column names are the values of **landuse_uri** raster or "landuse.txt" file. The Kc values can be determined as referenced by FAO (http://www.fao.org/docrep/x0490e/x0490e0b.htm).

A script “Kc_generator.py” in folder “related materials” is provided to assist generating the “Kc_d.txt” file for users prefer Python script method. Users can also prepare it with Microsoft Excel.

(9) The maximum value of initial abstraction coefficient (λ_{MAX}) parameter named “cIa_d.txt”, which provides the initial values of parameter λ_{MAX} for each day of a year. The initial values would be calibrated by multiplying an adjustment factor, c_Ia_m, a parameter that will be described in section 3.2 of this manual. The row index of the “cIa_d.txt” indicates the day of year. The values under column “c_Ia” are the initial values of parameter λ_{MAX}. The daily variation of initial λ_{MAX} accounts for the change of vegetation interception. As for the AXL example, the initial λ_{MAX} for the growing season is set to be 0.12 and 0.005 for the inactive/fallow season (Variable **cIa_d_File**, line 83)

2.2 Input Parameters

The parameters that need for DHM-WM local water balance routine are set in the model script directly, and include:

(1) The simulation periods

The simulation period is defined by the start day (variable **startDay**, line 74) and end day (variable **endDay**, line 75). Both parameters are in the format of “YYYY-M(M)-D(D)” in double quotation marks. For model calibration, the simulation period should cover both warm-up period and calibration period. Streamflow results need to be reported only for the calibration period, so routing analysis can be done only for the calibration period to save computing time. The start day of routing analysis (or the start day of calibration period) is set by users in variable **dayRouting** (line 90), which is also in the format of "YYYY-M(M)-D(D)" in double quotation marks.

(2) Calibrated parameters

There are at most 7 parameters that recommend calibration in DHM-WM local water balance routine (line 92-101), they are listed in Table 2-1.

Table 2-1. Summary of Calibrated parameters of DHM-WM local water balance routine

Variable in script	Parameter in theoretical document	Description	Range
CN_m*	S_{abs_mx}	Maximum value of absolute retention parameter (mm)	>0
c_Ia_m*	λ_{MAX}	Maximum value of initial abstraction coefficient	0-0.30
fc0	f_{c0}	Gravitational infiltration rate coefficient	0-1.000
Kb	K_b	Storage coefficient of base flow (d)	>0.5
tLag	g_{drain}	Lag coefficient for tile flow (h)	>0
tDrain	t_{drain}	Time required to drain water in soils with Hydrologic Soil Group D (h)	>0
CN_tile	CN_{tile}	Adjustment factor for CN values for tile-drained fields	0-1.00

If tile drainage is not important in the study watershed, variable **tile** (line 96) should be set to be False, then only the first 4 parameters in Table 2-1 needs calibration. Note: the first two calibrated parameters (variable **CN_m** in line 92 and variable **c_Ia_m** in line 93) are adjustment factors for parameter S_{abs_mx} and parameter λ_{MAX}. The initial value of λ_{MAX} is set in the input text file "cIa_d.txt" described above. The initial values of S_{abs_mx} are associated with curve numbers (CN) as: $S_{abs\_mx\_default}$ = 2.281*(25400/CN-254), and the initial CN values are set in land use related property file "landuse.txt" described above.

(3) Tile drainage related parameters (optional)

If tile flow module is active (the variable **tile** in line 96 is set to be True), then there are some tile drainage related parameters need to be set by users, including: variable **tLag** (line 99), variable **tDrain** (line 100) and variable **CN_tile** (line 101) in Table 2-1 that needs calibration; the depth of tiles to surface in mm (variable **tileDepth** in line 102), the maximum tile flow that the tiles can drain in mm/m^2/day (variable **tFlow_mx** in line 103), and the list defining the land use types where tile drainage systems are applied (variable **tileList**, line 104). Parameters **tileDepth** and **tFlow_mx** are recommended to be determined by survey rather than calibration. The value of **tileList** is the land use names (aligned with those set in the "landuse.txt" file) separated by commas and all in a pair of square brackets [].

(4) Variables that need initiation

There is one variable in DHM-WM local water balance routine that needs users to initiate and is updated daily during simulation, thus becoming representative to the actual condition after warm-up period. That is, variable **TmpSoil** (line 106) represents the average soil temperature in degrees Celsius, its value should include one digit after decimal point.

(5) Other parameters

There are some other parameters need by DHM-WM local water balance routine that do not need calibration but also need to be set by users. Variable **Latitude** (line 108) is the geographic latitude of the center of the study watershed, the value is in decimal degrees and only one digit after decimal point is enough. Variable **soil_bd** (line 109) represents the watershed average soil bulk density, in Mg/m^3. Variable **n_Mnng_streams** (line 110) indicates the Manning's n roughness coefficient for streams/channels. Variable **v_lake** (line 111) is the velocity of water flow in lakes/reservoirs/ponds in m/s. If **v_lake** is not easy to determine, use the default value.

2.3 Set output text file names and run the model

The outputs of DHM-WM include rasters of surface runoff depth and travel time of surface runoff, as well as text files recording time series. The raster output data are in GeoTiff format and stored in the folder defined by variable **outputGdb** (line 30). The text files are stored in the folder defined by variable **outputTxt** (line 32) and the file names are set in line 114-119 of the script, including: some important parameters that will help understand the model performance (variable **parameter_Output**, line 114), surface flow at the specified discharge sites (variable **Qsur_Output**, line 115), ground water flow at the specified discharge sites (variable **Qgrd_Output**, line 116), total streamflow at the specified discharge sites (variable **Qtot_Output**, line 117), and tile flow at the specified discharge sites if tile flow module is active (variable **Qtile_Output**, line 119).

Now, users can run the model by clicking “Run Module” below “Run” in the menu, or by pressing F5. Then, a screen will prompt out, showing model processing. The computing time depends on the watershed size, the grid cell size, the number of target sites selected to report, and the computer properties. For a watershed of 43 km^2 with a resolution of 10 m, one site to report hydrographs in the AXL example, DHM-WM simulation runs take about 10 minutes per simulation year on a desktop with an Intel(R) Core(TM) i5 3.30 GHz CPU and 8.00 GB RAM. When the simulation ends, the screen will prompt out "Voila! DHM-WM local water balance routine has completed your task." Then, you can check the output results.

2.4 Output Data

The outputs of DHM-WM include raster of surface runoff depth and travel time of surface runoff, as well as text files of simulated important parameters, flow components (surface flow, ground water flow, total streamflow, and tile flow if tile module is active) at the specified discharge sites. All these data are reported on a daily time step.

2.4.1 Output raster

The output raster data are illustrated by Figure 2-1 as an example.

(1) Distributed surface runoff depth in mm which is named as "qYYYYMMDD".

(2) Distributed travel time of surface runoff across a grid cell in seconds which is named as "tYYYYMMDD". Note: You can find the grid cells with no surface runoff have a very long travel time, which is meaningless but just for calculation needs.

(3) Distributed cumulative travel time of surface runoff to the watershed outlet in hours which is named as "tConYYMMDD". Note: Similar to the "tYYYYMMDD" data, the grid cells with no surface runoff but a great value of cumulative travel time is meaningless.

(4) Spatial grids of some variables for users to do insight check of the model performance, including: “S_abs_mx”, the parameter Sabs_mx distribution used in this simulation; “soil_HG”, the distributed soil hydrologic groups (value 1 refers to group A, 2 for B, 3 for C and 4 for D); “soil_k”, the distributed average saturated hydraulic conductivity of soil, in m/day; “soil_z”, the distributed soil depth to impervious/restrictive layer in meters; “soil_Pe”, the distributed effective porosity of soil; and “soil_Pt”, the distributed total porosity of soil.

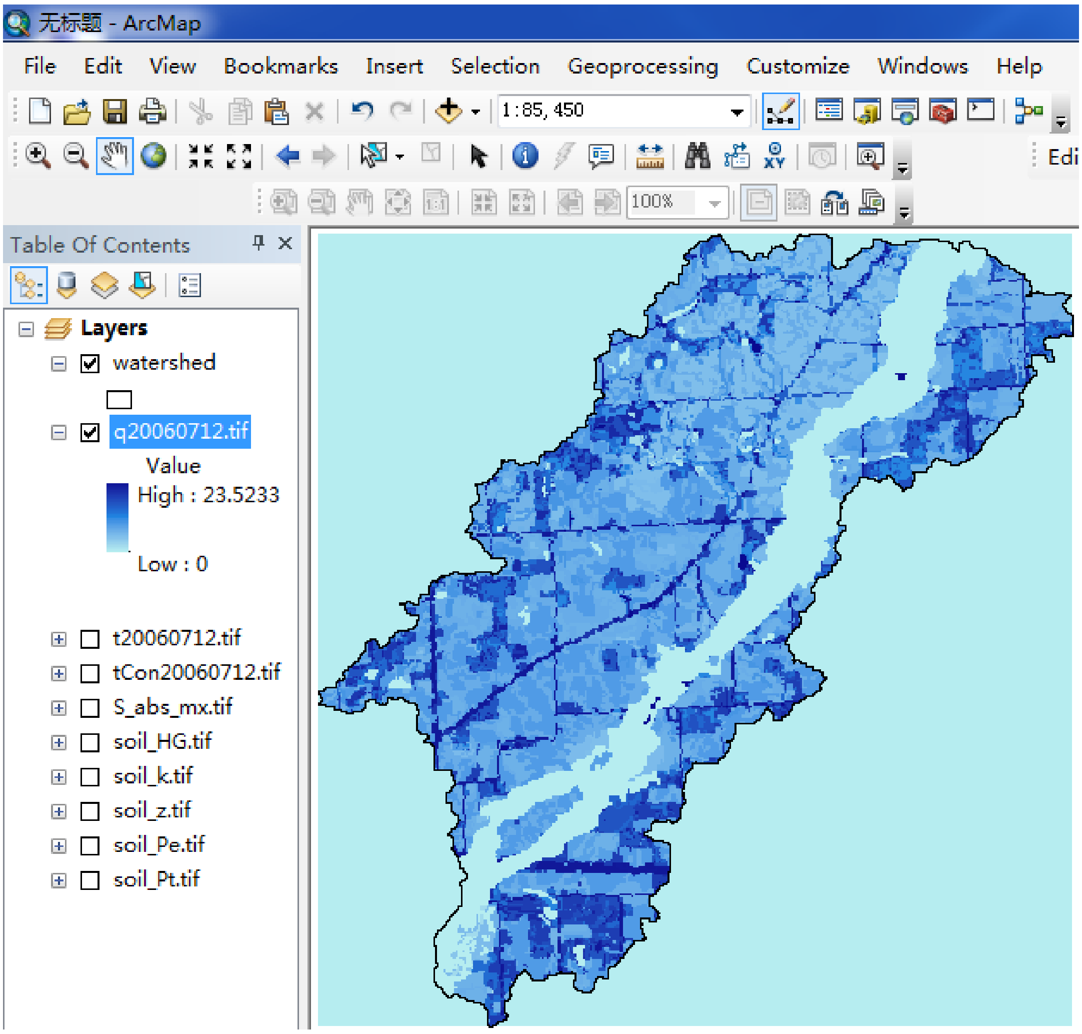


Figure 2-1. Example of output raster data

2.4.2 Output text files and results statistics for model evaluation

The output text files are named by users (see section 2.3). The column names of the output flow (surface flow, groundwater flow, total streamflow, and tile flow if tile flow module is active) files are generated from the "name" field of "discharge_sites.shp" file (variable **sitesDischarge**), representing different specified discharge sites. The values are in cubic meters per second (m^3/s).

The total streamflow time series can be used to compare with the observed data, and calculate statistic indicators to evaluate model performance. DHM-WM provides a script "resultsStats.py" stored in the subfolder "results statistics" to automatically calculate percent bias (P_{bias}), coefficient of determination (R^2) and Nash-Sutcliffe coefficient (Ens) of total streamflow. You can copy the simulated total streamflow file ("fOutput_test.txt" as in the AXL example) to the subfolder "results statistics". Also, you need to prepare a file stores the observed total streamflow data as the "flow_obs.txt" file in the same subfolder. The column names of the "flow_obs.txt" file should be the same with the simulated "fOutput_test.txt" file. Users need to set the 4 variables of the script "resultsStats.py" before running it: **Sim_file** (the name of the simulated data), **Obs_file** (the name of the observed data), **startDay** (the start day of the statistic period) and **endDay** (the end day of the statistic period).

3. DHM-WM Global Water Balance Routine

3.1 Input Data

The input data for DHM-WM global water balance routine include spatial data and text files, which are stored separately in two folders. Users should set the path of a folder that spatial data are stored (in a variable **inputFolder**, line 26); and the path of a folder that text files are stored (in a variable **python_inputFd**, line 28).

3.1.1 Input spatial data

The spatial data that users need to input include: raster data in GeoTiff format (flow direction, soil, land use, DEM, and gridded precipitation data if available); and vector data in ArcGIS Shapefile format (delineated watershed, and delineated streams from watershed preparation analysis, reference points for reference zone delineation, output sites for discharge simulation, output sites for soil moisture simulation). Raster data should be stored in a subfolder named “rasters” within the **inputFolder**; and vector data should be stored in a subfolder named “vectors” within the **inputFolder**. These data are described in detail below.

(1) Flow direction data named “fdr.tif”, which is the output data of watershed preparation and should be stored just in the “rasters” subfolder. (Variable **fdr_uri**, line 47)

(2) Soil data named “soil.tif”, which is the spatial distribution of the soil type. The “value” field of the raster will be used to connect to their properties data which are prepared in a text file. “soil.tif” should cover the extent of the delineated study watershed. (Variable **soil_uri**,

line 48)

(3) Land use data named “landuse.tif”, which is the land use/land cover raster of the study watershed. The “value” field of the raster will be used to connect to their properties or parameters which are prepared in a text file. The “landuse.tif” data should cover the extent of the delineated study watershed. (Variable **landuse_uri**, line 49)

(4) DEM data named “dem.tif”, which is the same as the input data of “watershed preparation” analysis. The elevation unit should be meters. (Variable **dem_uri**, line 50)

(5) Delineated watershed vector from watershed preparation analysis named “watershed.shp”, which is the output data from the watershed preparation. (Variable **watershed_uri**, line 51)

(6) Delineated stream vector from watershed preparation analysis named “stream_DL.shp”, which contains a field named “strWidth” that records the stream width data in meters. (Variable **stream_uri**, line 52)

(7) Reference points named “reference_points.shp”, which is used to delineate reference zones to calculate STI, see “DHM-WM theoretical document” section 2.2.3. The “reference_points” are the drainage outlets of reference zones, prepared by users based on a principle that each significant lake, pond and stream reach drains a separate reference zone. The points should either lie on the polylines of “stream_DL” data or on the flowlines draining ponds/lakes/researvoirs. (Variable **ref_point_uri**, line 53)

(8) Output sites for discharge simulation named “discharge_sites.shp”, which are the target

sites that need to report discharge time series. Generally, the more the number of target sites, the longer it will take for model computation. The points in “discharge_sites” data should lie on the watercourse polylines of “stream_DL.shp” data. Each point in the “discharge_sites.shp” should have a field “name” which describe the name of each site, the simulating results will be reported using these names. (Variable **sitesDischarge_uri**, line 54)

In addition, the watershed outlet should be one of the points in “discharge_sites.shp”. The name of the watershed outlet is provided in Variable **outlet_name** in line 55 (as “I” in the example)

(9) Output sites for soil moisture simulation named “moisture_sites.shp”, which are the target sites that need to report soil moisture time series. The “moisture_sites.shp” should have 3 fields: “name” that defines the name of each moisture sites, “landuse” and “soil” which describe the land use type and soil type of each moisture sites, the values of fields “landuse” and “soil” should be aligned with the values in “landuse.tif” and “soil.tif” raster data. (Variable **sitesMoisture_uri**, line 56)

(10) Gridded precipitation data, which are optional. If gridded precipitation data are available, set the variable **haveDisPrcp** (line 58) to be True, and set the name of the folder where gridded precipitation data stores (Variable **DisPrcpFolder**, line 60). The folder should be within the sub-folder “rasters” of **inputFolder**, and the default name is “prcp/”. The gridded precipitation data should be named as “prcpYYYYMMDD.tif” such as “prcp20030101.tif” (see LREW_example). The gridded precipitation data are interpolated from monitoring data at precipitation gauges by Inverse Distance Weighting (IDW) method. The gauge data are

prepared in ArcGIS .xyz format and stored in the folder “data_input\grids\prcp\point data”. In the same folder, there is a script “precip_point2grid.py” that can do the interpolation.

3.1.2 Input text files

The text files that need for DHM-WM global water balance routine are stored in the folder defined by variable **python_InputFd** (line 28), and include:

(1) Land use related properties and parameters named “landuse.txt”, which include columns named “Value”, “Name”, “Manning_n”, “CN_A”, “CN_B”, “CN_C”, “CN_D”. “Value” column is used to connect to the “Value” field of “landuse” raster data. “Name” column is a 4-character upper case symbol to imply the land use type, common “Name”s are listed in Table 1. “CN_A”, “CN_B”, “CN_C”, “CN_D” are the default curve numbers for Hydrologic Soil Group (HSG) type A, B, C, D, respectively. The default curve numbers are determined based on NRCS National Engineering Handbook (NEH) 630 table 9, see the file in the folder of “related materials”. (Variable **landuseFile**, line 64)

(2) Soil type related properties named “soil.txt”, which include columns named “Value”, “HSG”, “Soil_z(m)”, “Soil_k(m/day)”, “Pe”, “Pt”. “Value” column is used to connect to the “Value” field of “soil” raster data. “HSG” is the HSG type defined by NRCS, which can be directly extracted from SSURGO database. Note: the values of “HSG” columns should be numbers, “HSG” value 1 refers to type A, value 2 refers to type B, value 3 refers to type C, value 4 refers to type D. “Soil_z(m)” is the soil depth to impervious/restrictive layer (in the unit of meters), which is determined by checking the saturated hydraulic conductivity with depth.

Impervious/restrictive layer is defined as one with conductivity below 10 mm h^{-1}. "Soil_k(m/day)" is the average saturated hydraulic conductivity of the pervious soil profile (in the unit of meters per day), which is the weighted average saturated hydraulic conductivity of different layers above impervious/restrictive layer. "Pe" and "Pt" are the average effective porosity and total porosity of the pervious soil profile, which can be determined with other soil properties, such as sand-silt-clay percentages. Two methods are recommended to estimate "Pe" and "Pt" (see "Porosity estimation" in "related materials"): (a) a simple porosity calculator based on a reference file from Argonne laboratory, which need only sand-silt-clay percentages; (b) a soil water characteristics tool of SPAW hydrology developed by USDA ARS incorporated with Washington State University (https://hrsl.ba.ars.usda.gov/SPAW/Index.htm), which needs sand-silt-clay percentages, organic matter content, salinity and compaction. (Variable **soilFile**, line 65)

(3) Daily precipitation data named "prcp_d.txt", which contains two columns: "date" and "prcp_d(mm)". Column "date" is in the format of "YYYY-M(M)-D(D)" and should cover all simulation periods, including warm-up period, calibration period and validation period. Do not contain missing values, fill it as reasonable as possible. "prcp_d(mm)" is the daily precipitation in mm. (Variable **prcp_File**, line 67)

(4) 12-h and 24-h synthetic rainfall distribution data named "12raindis.txt" and "24raindis.txt", both of which contain two columns: "hour" and "cmmlPrcp". These data are used to estimate rainfall distribution within a day and calculate the duration of effective rainfall or runoff generation: use "12raindis.txt" when daily precipitation is below 50 mm and

"24raindis.txt" when daily precipitation is above 50 mm. "cmmlPrcp" represents the cumulative precipitation rate at a specific time "hour". Synthetic rainfall distribution can be estimated with rainfall frequency data using a spreadsheet tool provided by USDA NRCS ("Rainfall_Distribution_Spreadsheet_Tool.xlsx" in the folder of "related materials"). The sheet "Readme" in this tool provides links to information on rainfall, frequency and distribution, and where frequency data can be downloaded. "12raindis.txt" and "24raindis.txt" can be prepared with 3 steps. Step 1, calculate synthetic rainfall distribution with frequency data. Input rainfall frequency data into the "rainfall – in" column (C8-C17) in Sheet 1. Step 2, transform duration from 24-h to 12-h. Copy the estimated "Ratio" column (B45-B285) in Sheet 1 to "12-hour transformation" (column K) in sheet "duration transformation", note: Select the proper rainfall distribution type by clicking specific type in row 2, and paste the "Ratio" in proper place. Step 3, transform results from 0.1h increment to 0.5h increment. Copy the estimated "12-hour distr." (column N, in pink color) to the orange area in sheet "increment transformation", and the columns D-E are the 12-hour rainfall distribution that you can copy to the "12raindis.txt" file; copy the 24-h "Ratio" column (B45-B285) in Sheet 1 that you have copied in step 2 again and paste to the red area (Column I) in sheet "increment transformation", and the columns K-L are the 24-hour rainfall distribution that you can copy to the "24raindis.txt" file. (Variable **rainDis12_File** in line 68, and **rainDis24_File** in line 69)

(5) Daily temperature data named "tmp_d.txt", which contains columns named "date", "TEM_max", "TEM_min". Column "date" is in the format of "YYYY-M(M)-D(D)" and should cover all simulation periods, including warm-up period, calibration period and validation

period. Columns “TEM_max”, “TEM_min” represent maximum air temperature, minimum air temperature, respectively, in degrees Celsius. Missing values are represented as -99.9. Several missing values are OK, but more than 7 consecutive missing values may result in bias. (Variable **tmp_File**, line 70)

(6) Daily potential evapotranspiration (PET) data named “PET_d.txt”, which is optional. If observed PET data is available, set variable **haveDailyPET** to be “True” (line 71), and prepare “PET_d.txt” (Variable **PET_File**, line 73) in folder **python_inputFd.** There should be two columns in “PET_d.txt” file: “date” and “PET(mm)”. Column “date” is in the format of “YYYY-M(M)-D(D)” and should cover all simulation periods, including warm-up period, calibration period and validation period. Column “PET(mm)” is the PET of specific day in the unit of mm. Missing values are allowed in “PET_d.txt” file and the value of “PET(mm)” should be set to “-6999”. If observed PET data is unavailable, users need to set variable **haveDailyPET** to be “False”, and set the method used to calculate daily PET (variable **PET_method** in line 73). DHM-WM provided two methods: 1 represents Hargreaves method and 2 represents FAO Penman-Monteith method.

(7) Daily wind speed, sunlight duration, and humidity data, if **PET_method** is set to be 2 (FAO Penman Monteith). These data are needed for PET calculation using FAO Penman Monteith method. The wind speed data file named “wind_d.txt” (variable **win_File** in line 77) should contain at least two columns: “date” and “WIN_ave” (daily average wind speed measured at 10 m height, in m/s), missing values are represented as -99.9; the sunlight duration data file named “SSD_d.txt” (variable **SSD_File** in line 78) should contain two columns: “date” and

"SSD" (daily actual sunlight duration, in hr), missing values are represented as -99.9; the humidity data file named "RHU_d.txt" (variable **humidity_File** in line 79) should contain at least two columns, "date" and "RHU_ave" (daily average relative humidity, in %), missing values are represented as -99.

(8) Daily crop coefficients (Kc) of a year for different land covers named "Kc_d.txt", which is optional. If the user would like to account for the effect of vegetation on PET, daily crop coefficients (Kc) would be used to adjust PET for different land cover and different day. In this case, set variable **haveDailyKc** (line 81) to be True and prepare the "Kc_d.txt" data (variable **Kc_d_File**, line 83). The "Kc_d.txt" file has 366 rows representing the day of year. The number of columns is the number of land use types of the study watershed, and column names are the values of **landuse_uri** raster or "landuse.txt" file. The Kc values can be determined as referenced by FAO (http://www.fao.org/docrep/x0490e/x0490e0b.htm). A script "Kc_generator.py" in folder "related materials" is provided to assist generating the "Kc_d.txt" file for users prefer Python script method. Users can also prepare it with Microsoft Excel.

(9) Daily root depth (RD) of a year for different land covers named "RD_d.txt", which has 366 rows representing the day of year. The number of columns is the number of land use types of the study watershed, and column names are the values of **landuse_uri** raster data or "landuse.txt" file. The root depths for trees (forests, orchards) are set to be 10.00 meters (a value usually larger than the soil depths) all around the year; the root depth for grassland is set to be 0.30 meter all around the year; the root depth for lands without vegetation (urban,

residential, lakes, ponds, reservoirs, and fallow fields) is set to be 0 all around the year; the root depth for cultivated agricultural lands are defined by users. A script "RD_generator.py" in folder "related materials" is provided to assist generating the "RD_d.txt" file for users prefer Python script method. Users can also prepare it with Microsoft Excel. (Variable **RD_d_File**, line 84)

(10) The maximum value of initial abstraction coefficient (λ_{MAX}) parameter named "cIa_d.txt", which provides the initial values of parameter λ_{MAX} for each day of a year. The initial values would be calibrated by multiplying an adjustment factor, c_Ia_m, a parameter that will be described in section 3.2 of this manual. The row index of the "cIa_d.txt" indicates the day of year. The values under column "c_Ia" are the initial values of parameter λ_{MAX}. The daily variation of initial λ_{MAX} accounts for the change of vegetation interception. As for the LREW example, the initial λ_{MAX} for the growing season is set to be 0.16 and 0.12 for the inactive/fallow season because in the watershed forests are most deciduous and agricultural cultivation is generally for one season. If there is no significant change of vegetation coverage of the study watershed, the initial λ_{MAX} would be set to be constant through a year. (Variable **cIa_d_File**, line 86)

3.2 Input Parameters

The parameters that need for DHM-WM global water balance routine are set in the model script directly, and include:

(1) The simulation periods

The simulation period is defined by the start day (variable **startDay**, line 90) and end day (variable **endDay**, line 91). Both parameters are in the format of “YYYY-M(M)-D(D)” in double quotation marks. For model calibration, the simulation period should cover both warm-up period and calibration period. Streamflow results need to be reported only for the calibration period, so routing analysis can be done only for the calibration period to save computing time. The start day of routing analysis (or the start day of calibration period) is set by users in variable **dayRouting** (line 93), which is also in the format of “YYYY-M(M)-D(D)” in double quotation marks.

(2) Calibrated parameters

There are at most 8 parameters that recommend calibration in DHM-WM global water balance routine (line 95-102), they are listed in Table 3-1. If irrigation is not applied in the study watershed, parameter SM_{irrg} (variable **M_irrg**) should be set to be 0. Note: the first two calibrated parameters (variable **CN_m** and variable **c_Ia_m**) are adjustment factors for parameter S_{abs_mx} and parameter λ_{MAX}. The initial value of λ_{MAX} is set in the input text file “cIa_d.txt” described above. The initial values of S_{abs_mx} are associated with curve numbers (CN) as: $S_{abs\_mx\_default}$ = 2.281*(25400/CN-254), and the initial CN values are set in land use related property file “landuse.txt” described above.

Table 3-1. Summary of Calibrated parameters of DHM-WM global water balance routine

Variable in script	Parameter in theoretical document	Description	Range
CN_m*	S_{abs_mx}	Maximum value of absolute retention parameter (mm)	>0
c_Ia_m*	λ_{MAX}	Maximum value of initial abstraction coefficient	0-0.30
fc0	f_{c0}	Gravitational infiltration rate coefficient	0-1.000
Kb	K_b	Storage coefficient of base flow (d)	>0.5
soil_f0	f_{soil}	Exponential decline coefficient of conductivity with soil depth (m^{-1})	1.0-20.0
K_sf	k_f	Adjustment parameter for f_{soil} on saturated condition	0-1.00
tM0	$\overline{tM}$	Watershed average time to drain for moisture loss estimation (h)	0-200
M_irrg	SM_{irrg}	Relative soil moisture level below which irrigation is applied	0-1.000

(3) Irrigation related parameters (optional)

If irrigation module is active (the variable **M_irrg** in line 102 is set to be greater than 0), then there are some irrigation related parameters need to be set by users, including: the start day of year (variable **irrg_start**, line 105) and end day of year (variable **irrg_end**, line 106) defining the period for irrigation (usually growing season of crops), and the list defining the land use types that apply irrigation (variable **irrgList**, line 107). The value of **irrgList** is the land use names (aligned with those set in the "landuse.txt" file) separated by comma and all in a pair of square brackets [].

(4) Variables that need initiation

There are two variables in DHM-WM global water balance routine that need users to initiate and is updated daily during simulation, thus becoming representative to the actual condition after warm-up period. Variable **res** (line 109) indicates the water table depth in water bodies in mm ($\overline{D}$ in equation (7) of the theoretical document), its value should have one digit after decimal point. Variable **TmpSoil** (line 110) represents the average soil temperature in degrees Celsius, also include one digit after decimal point for its value.

(5) Other parameters

There are some other parameters need by DHM-WM global water balance routine that do not need calibration. Variable **Latitude** (line 112) is the geographic latitude of the center of the study watershed, the value is in decimal degrees and only one digit after decimal point is enough. Variable **soil_bd** (line 113) represents the watershed average soil bulk density, in Mg/m^3. Variable **n_Mnng_streams** (line 114) indicates the Manning's n roughness coefficient for streams/channels. Variable **v_lake** (line 115) is the velocity of water flow in lakes/reservoirs/ponds in m/s. If **v_lake** is not easy to determine, use the default value.

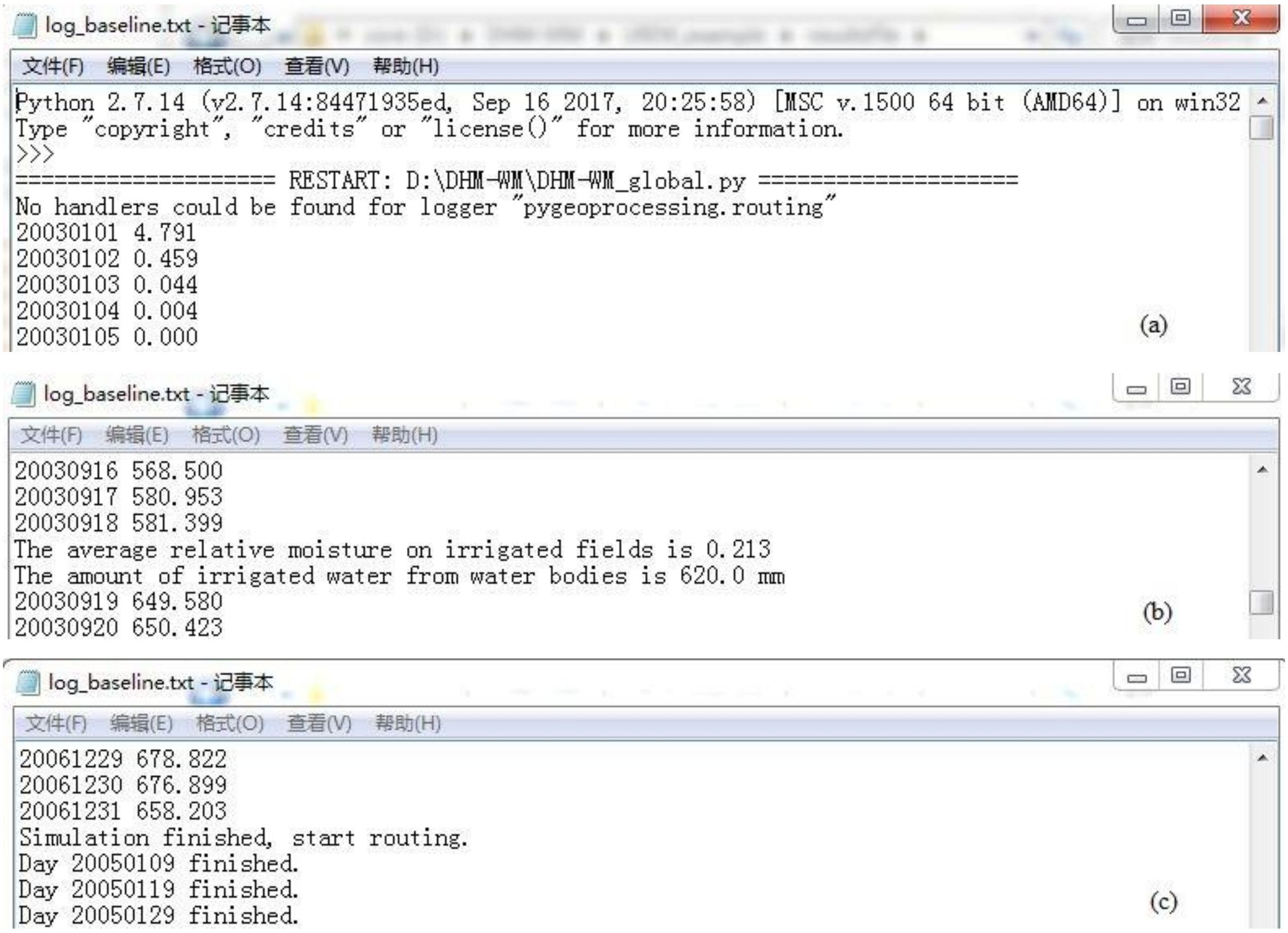
3.3 Set output text file names and run the model

The outputs of DHM-WM include rasters of surface runoff depth and travel time of surface runoff, as well as text files recording time series. The raster output data are in GeoTiff format and stored in the folder defined by variable **outputGdb** (line 30). The text files are stored in the folder defined by variable **outputTxt** (line 32) and the file names are set in line 118-124 of the script, including: some important parameters that will help understand the model performance (variable **parameter_Output**, line 118), surface flow at the specified discharge sites (variable

Qsur_Output, line 119), ground water flow at the specified discharge sites (variable **Qgrd_Output**, line 120), subsurface return flow at the specified discharge sites (variable **Qrtn_Output**, line 121), total streamflow at the specified discharge sites (variable **Qtot_Output**, line 122), relative moisture content of the unsaturated layer at the specified moisture sites (variable **M_Output**, line 123), and subsurface water table depth at the specified moisture sites (variable **H_Output**, line 124).

Now, users can run the model by clicking “Run Module” below “Run” in the menu, or by pressing F5. Then, a screen will prompt out, showing model processing. The computing time depends on the watershed size, the grid cell size, the number of target sites selected to report, and the computer properties. For a watershed of 50 km^2 with a resolution of 30 m, three sites to report hydrographs and five sites for soil moisture in the LREW example, DHM-WM simulation runs take about 2 minutes per simulation year on a desktop with an Intel(R) Core(TM) i5 3.30 GHz CPU and 8.00 GB RAM. When the simulation ends, the screen will prompt out “Voila! DHM-WM global water balance routine has completed your task.” Then, you can check the output results. But before that, it is suggested to save the simulation processing screen as a log file by clicking “Save as” below “File” button in the menu. In the LREW example, the file is named as “log_baseline.txt”, see Figure 3-2.

At the start of the log file (Figure 3-2(a)), the warning “No handlers could be found for logger “pygeoprocessing.routing” can be safely ignored.


```
log_baseline.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
Python 2.7.14 (v2.7.14:84471935ed, Sep 16 2017, 20:25:58) [MSC v.1500 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
==================== RESTART: D:\DHM-WM\DHM-WM_global.py ====================
No handlers could be found for logger "pygeoprocessing.routing"
20030101 4.791
20030102 0.459
20030103 0.044
20030104 0.004
20030105 0.000
(a)

log_baseline.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
20030916 568.500
20030917 580.953
20030918 581.399
The average relative moisture on irrigated fields is 0.213
The amount of irrigated water from water bodies is 620.0 mm
20030919 649.580
20030920 650.423
(b)

log_baseline.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
20061229 678.822
20061230 676.899
20061231 658.203
Simulation finished, start routing.
Day 20050109 finished.
Day 20050119 finished.
Day 20050129 finished.
(c)
```

Figure 3-2. Example of model processing file

Then, it writes the day of simulation in the format of “YYYYMMDD” and the solved water table depth (variable **res** in the script, or $\overline{D}$ in equation (7) of the theoretical document) of the simulation day in mm. Sometimes, there are two sentences between the water table depth (Figure 3-2(b)), which means irrigation was simulated to be applied on the day below it. It first tells the average relative moisture on irrigated fields before irrigation, the value must be smaller than the parameter SM_{irrg} defined by variable **M_irrg**, then it tells the amount of irrigated water extracted from water bodies simulated by DHM-WM. After each day in the specified simulation period is simulated, the model starts to do routing analysis from the specified day defined by variable **dayRouting** (Figure 3-2(c)). The routing processing is reported on a ten-day interval.

3.4 Output Data

The outputs of DHM-WM include raster data of surface runoff depth and travel time of surface runoff, as well as text files of simulated important parameters, flow components (surface flow, subsurface return flow, ground water flow, total streamflow) at the specified discharge sites, relative moisture content of the unsaturated layer and water table depth at the specified moisture sites. All these data are reported on a daily time step.

3.4.1 Output raster data

The output raster data are illustrated by Figure 3-3 as an example.

(1) Distributed surface runoff depth in mm which is named as “qYYYYMMDD”.

(2) Distributed travel time of surface runoff across a grid cell in seconds which is named as “tYYYYMMDD”. Note: You can find the grid cells with no surface runoff have a very long travel time, which is meaningless but just for calculation needs.

(3) Distributed cumulative travel time of surface runoff to the watershed outlet in hours which is named as “tConYYMMDD”. Note: Similar to the “tYYYYMMDD” data, the grid cells with no surface runoff but a great value of cumulative travel time is meaningless.

(4) Spatial raster data of some variables for users to do insight check of the model performance, including: “S_abs_mx”, the parameter Sabs_mx distribution used in this simulation; “soil_HG”, the distributed soil hydrologic groups (value 1 refers to group A, 2 for B, 3 for C and 4 for D); “soil_k”, the distributed average saturated hydraulic conductivity of soil, in m/day; “soil_z”, the distributed soil depth to

impervious/restrictive layer in meters; “soil_Pe”, the distributed effective porosity of soil; “soil_Pt”, the distributed total porosity of soil; “STI”, the distributed Soil Topographic Index; “STI_ref”, the referenced STI for each drainage reference zone ($\overline{\mathrm{STI}}_{\mathrm{ref}}$ in equation (6) of the theoretical document); and “STI_rel”, the relative STI distribution, estimated as STI_rel = STI_ref – STI. A smaller value of STI_rel indicates wetter soils since water tends to concentrate rather than flow away.

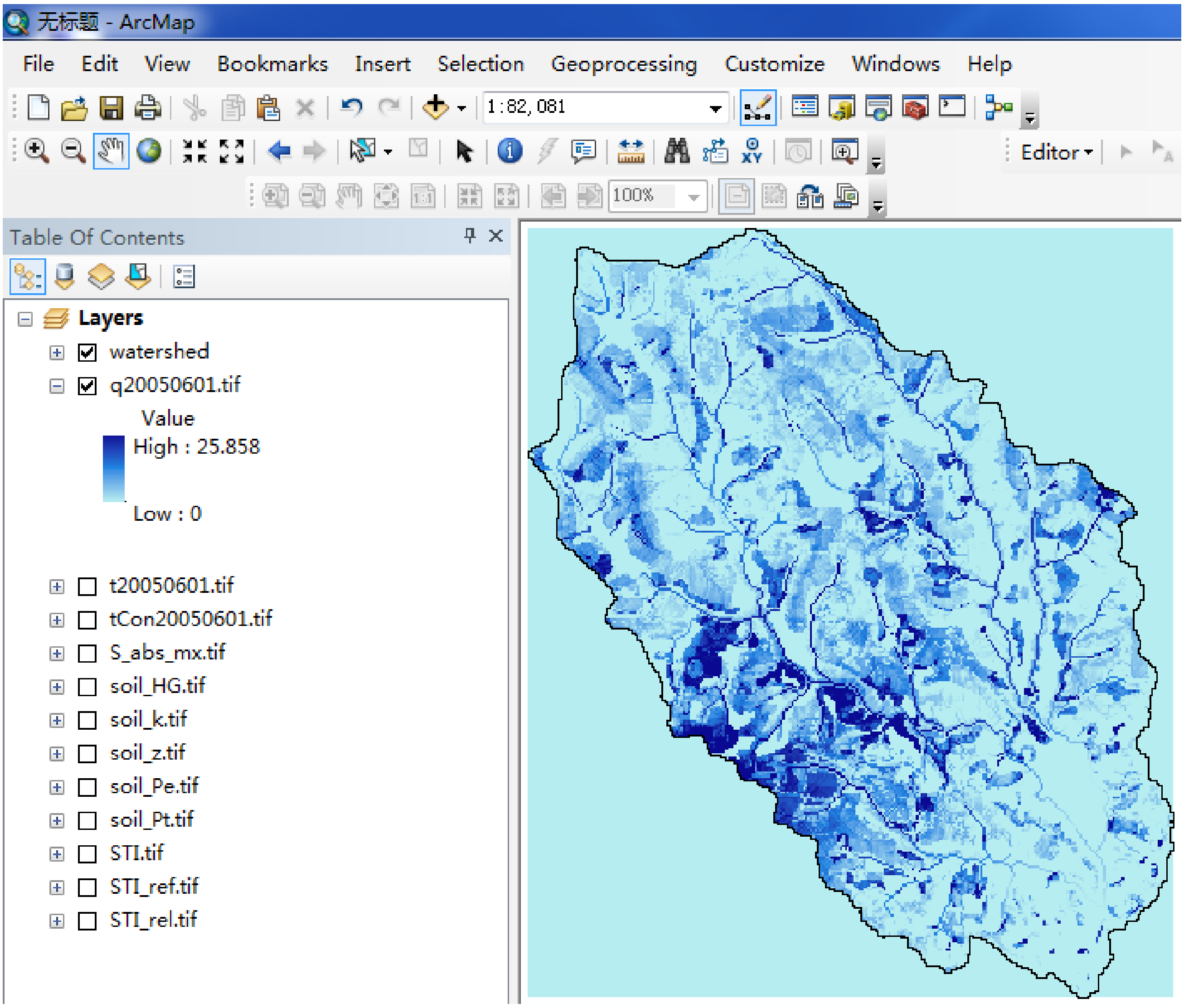


Figure 3-3. Example of output raster data

3.4.2 Output text files and results statistics for model evaluation

The output text files are named by users (see section 3.3). The column names of the output flow (surface flow, groundwater flow, subsurface return flow and total streamflow) files are generated from the "name" field of "discharge_sites.shp" file (variable **sitesDischarge**), representing different specified discharge sites. The values of flows are in cubic meters per second (m^3/s). The column names of the output relative moisture and water table depth files are generated from the "name" field of "moisture_sites.shp" file (variable **sitesMoisture**), representing different specified moisture sites. The soil moisture are relative water content.

The total streamflow time series can be used to compare with the observed data, and calculate statistic indicators to evaluate model performance. DHM-WM provides a script "resultsStats.py" stored in the subfolder "results statistics" to automatically calculate percent bias (P_{bias}), coefficient of determination (R^2) and Nash-Sutcliffe coefficient (Ens) of total streamflow. You can copy the simulated total streamflow file ("fOutput_baseline.txt" as in the LREW example) to the subfolder "results statistics". Also, you need to prepare a file stores the observed total streamflow data as the "flow_obs.txt" file in the same subfolder. The column names of the "flow_obs.txt" file should be the same with the simulated "fOutput_test.txt" file. Users need to set the 4 variables of the script "resultsStats.py" before running it: **Sim_file** (the name of the simulated data), **Obs_file** (the name of the observed data), **startDay** (the start day of the statistic period) and **endDay** (the end day of the statistic period).

Appendix

Table 1. Common land use names used in DHM-WM

4-character name	Land use type
FRST	Forest
RNGB	Shrubs
WETF	Riparian forested wetland
PAST	Grassland, pasture
URBR	Urban or residential land
WATR	Open waters (Lakes/ponds/reservoirs)
WETL	Wetland
AGRL	General agricultural field
RNGE	Fallow field
CORN	Corn field
SOYB	Soybean field
WWHT	Winter wheat field
ALFA	Alfalfa

Note: The land uses used in DHM-WM can be general classifications with "AGRL" representing all general agricultural field or specific classifications with different types for different crops. The decision depends on users. When users prepare their land use properties file, it is recommended that the above listed land use types are named as the listed 4-character names, users can define other names for other land use types.