**CACIE Tool #32** – **maxDose (maxDose.py)**

**Version** **1.0**

**QA**: **QA**

# Description and Purpose

This tool calculates the maximum dose per exposure pathway over user-specified time intervals and spatial extents. When invoked, maxDose reads a control file, consumes the output of the ca-dosecalc tool, and exports results as a collection of comma separated value (CSV) files that will be used in preparation of the CA report. The maximum dose is calculated for each time step, time interval, spatial domain, and exposure pathway (i.e. exposure route).

# Functional Requirements

The following are the functional requirements (FR) of maxDose

* This tool must:
  + FR-1: process a user-specified input control file.  *Additional requirements relating to the control file are specified below.*
  + FR-2: consume the output of the ca-dosecalc tool as input
  + FR-3: consume one or more [spatial] domain definition files
* The tool’s input control file must allow the user to:
  + FR-4: specify a directory where the output is stored. *Requirements relating to output files are described below.*
  + FR-5: define multiple time intervals
  + FR-6: define multiple domains
  + FR-7: define a domain by indicating a name and a path to the domain input file
  + FR-8: define a time interval by specifying a start year and end year in the control file.
* The tool must output:
  + FR-9: all results in CSV format
  + FR-10: the maximum dose results for each time interval specified in the control file.  *Requirements calculating the maximum dose results are described below.*
  + FR-11: the maximum dose results for each domain specified in the control file
  + FR-12: information relating to where and when the maximum dose occurred: the, year, row, column, and layer.
* The tool must compute the maximum dose by aggregating:
  + FR-13: For each time interval, domain, and exposure pathway, compute the maximum dose over the domain for each year in the interval.
  + FR-14: For each time interval, domain, and exposure pathway, compute the maximum dose over the domain.

# Software Requirements Specifications

This tool requires:

* python 3.6+, with the following third-party libraries:
  + pandas, version 1.0+

# Software Design Description

Arguments:

This tool accepts one input argument: a string indicating the path to the input control file.

Input Files:

**Input control file**

This file provides maxDose with important input, such as where to find the dose data. It must be in JavaScript Object Notation (JSON) format and have the following structure:

{

"copc":"U235",

"dosepath":"data/U235.csv",

"domains":[

{"name":"inner", "fpath":"innerDomainFile.csv"},

{"name":"outer", "fpath":"outerDomainFile.csv"},

{"name":"ca99", "fpath":"ca99DomainFile.csv"}

],

"outputdir":"output",

"dateranges":[

{"start\_year":2070, "end\_year":3070},

{"start\_year":3070, "end\_year":12070}

]

}

The field names are required and must be lower case. Users may specify additional fields, such as comments, but they are ignored by this tool.

* copc
  + this is a label that is added to the output filenames for easy identification
* dosepath
  + path to the CSV file containing the dose data produced by the **ca-dosecalc** tool.
* domains
  + A list of objects that define the name and location of the domain input file(s). Each entry in the list must be an object with the following fields (additional fields are ignored):
    - name
      * A label added to the relevant output filenames for easy identification
    - fpath
      * A path to the domain definition file. The structure of a domain definition file is defined below.
      * If an empty string, “”, is specified, the domain is assumed to be the entire model domain.
* outputdir
  + A path to an existing directory where the output will be stored
* dateranges
  + A list of objects that define the interval over which the maximum dose is calculated. Each entry in the list is an object with the following fields (additional fields are ignored).
    - start\_year
      * an integer defining the start year of the interval (inclusive)
    - end\_year
      * an integer defining the end year of the interval (inclusive)

**Domain definition file**(s)

A domain definition file identifies MODFLOW grid cells that this tool will include when calculating the maximum dose statistics. Cells are identified by their ROW and COLUMN index; those not included in the domain definition files are ignored by this tool.

In the practices, three domain files will typically be used: these define cells 1) on and outside the Inner Area Boundary, 2) on and outside the Outer Area Boundary, and 3) on and outside the CA-99 boundary.

The definition file must be a CSV file with the first line as a header. Additional rows must contain the row and column indices as integers, separated by a comma, “,” character. Row and column indices are assumed to be 1-indexed.

An example Domain Definition File:

row,column  
1,1  
1,2  
1,3

**Dose file**

The dose file is assumed to be taken directly from the output of the CA/CIE qualified tool **ca-dosecalc**. This file is a CSV file containing the calculated dose for each exposure pathway at every point in time and space in the model domain. Each row represents a unique space/time/pathway.

The first row is header text with column names. Columns correspond to:

|  |  |  |  |
| --- | --- | --- | --- |
| **Index** | **Column name** | **Type** | **Description** |
| 1 | elapsed\_tm | Integer | The number of days since the start of the model corresponding to this row’s data |
| 2 | model\_date | Text | The calendar date in YYYY-MM-DD format corresponding to elapsed\_tm. The calendar date is calculated as the model start date (determined by the modeldate input parameter) plus elapsed\_tm calendar days. Leap years are observed. |
| 3 | soil | Text | The soil type of the grid cell (row/col) |
| 4 | pathway | Text | The exposure pathway |
| 5 | cell\_row | Integer | grid index (1-based) |
| 6 | cell\_column | Integer | grid index (1-based) |
| 7 | cell\_layer | Integer | grid index (1-based) |
| 8 | concentration | Float | The concentration, in units of the MT3D UCN file, multiplied by the conversion parameter. |
| 9 | dose\_factor | Float | The dose factor for the pathway/soil type  Units are the same as those provided in dosefactsFile |
| 10 | Dose | Float | The calculated dose. Equivalent to column 8 \* column 9. |

Output Files:

The number of output files of this tool depends on the number of time intervals and spatial domains defined in the input control file. The form of the output mirrors the dose input file except the reported dose is the maximum for the given spatial domain and time interval. The given domain and interval are specified in the file’s name. There are two types of output file:

* **Maximum dose by pathway as a timeseries**. Files of this type are titled max\_for\_pathway\_for\_time\_[COPC]\_[DOMAIN]\_yr[START-END].csv. Each row in this file type reports the maximum over the spatial domain DOMAIN for the specified year and pathway. The remaining columns denote the properties of where and when the maximum dose occurred.
* **Maximum dose by pathway.** Files of this type are titled max\_for\_pathway\_[COPC]\_[DOMAIN]\_yr[START-END].csv. Each row reports the maximum dose for the given pathway over the domain and time interval. The remaining columns denote the time and location of where the maximum dose occurred.

Tool Runner:

The following is the shell script configuration that will be passed as an argument to the Tool Runner for qualified runs:

{directory path to repository}/pylib/camaxdose/maxDose.py {configuration file path/name}

Code Review:

Code walkthrough was performed by Neil Powers on 08/6/2020. No impacts to other repository tools or shared library dependencies were identified for the Max Dose tool

# Requirements Traceability Matrix

The requirements traceability matrix for the maxDose tool is presented in Table 1.

| Table 1  Requirements Traceability Matrix | | | |
| --- | --- | --- | --- |
| **Functional Requirement ID** | **Requirement Description** | **Test Case** | **Pass/Fail** |
| QA Level | Installation Test Case | CACIE-maxDose-IT-1 |  |
| FR-1 | Process a user-specified input control file. | CACIE- maxDose-AT-1 |  |
| FR-2 | Consume the output of the ca-dosecalc tool as input | CACIE- maxDose-AT-1 |  |
| FR-3 | Consume one or more [spatial] domain definition files | CACIE- maxDose-AT-1 |  |
| FR-4 | Specify a directory where the output is stored | CACIE- maxDose-AT-1 |  |
| FR-5 | Define multiple time intervals | CACIE- maxDose-AT-1 |  |
| FR-6 | Define multiple domains | CACIE- maxDose-AT-1 |  |
| FR-7 | Define a domain by indicating a name and a path to the domain input file | CACIE- maxDose-AT-1 |  |
| FR-8 | Define a time interval by specifying a start year and end year in the control file | CACIE- maxDose-AT-1 |  |
| FR-9 | All results in CSV format | CACIE- maxDose-AT-2 |  |
| FR-10 | The maximum dose results for each time interval specified in the control file | CACIE- maxDose-AT-2 |  |
| FR-11 | The maximum dose results for each domain specified in the control file | CACIE- maxDose-AT-2 |  |
| FR-12 | Information relating to where and when the maximum dose occurred: the, year, row, column, and layer | CACIE- maxDose-AT-2 |  |
| FR-13 | For each time interval, domain, and exposure pathway, compute the maximum dose over the domain for each year in the interval | CACIE- maxDose-AT-2 |  |
| FR-14 | For each time interval, domain, and exposure pathway, compute the maximum dose over the domain | CACIE- maxDose-AT-2 |  |

# Installation Test Plan and Acceptance Test Plan Cases

This tool will be tested on the same virtual machine that is used in production runs; therefore, acceptance tests 1 and 2 also serve as the installation test. The acceptance test plan cases are presented in Table 2 and 3.

| Table 2  **maxDose Acceptance Test Plan Case 1** | | | |
| --- | --- | --- | --- |
| **maxDose Acceptance Testing**  **CACIE-maxDose – AT-1** | | **Date:** | |
| **Tool Runner Log File Location for this test:**  **[PUT LINK TO THE DIRECTORY HERE]** | | **Test Performed By:** | |
| **Testing Directory: /home/ca/dose/test-maxDose** | | | |
| **Test Step** | **Test Instruction** | **Expected Result** | **Test Result  (Pass/Fail)** |
| You will need to log in to the server to perform this test. Please request the server location and login credentials from the lead developer. | | | |
| 1 | cd into the test directory by typing into the console and press enter:  cd ~/dose/test-maxDose |  | NA |
| 2 | Run the test script by typing into console and press enter:  ./testMaxDoseAT1.sh | The script begins to run and the console notifies you that Acceptance test 1 has started and that the output is logged to output/testlogfile.txt |  |
| 3 | Obtain information from the input control file and enter it into the fields in the box on the right:  Type the following into the console and press enter  vim -R inputs/testControlInput.json | 1) Number of unique objects in the “domain” field (each object has a “name” and “fpath” attribute):  NDOMAINS = \_\_\_\_\_\_\_\_\_\_\_\_  2) Number of unique objects in the “dateranges” field (each object has a “start\_year” and “end\_year” attributes):  **NINTERVALS** = \_\_\_\_\_\_\_\_  3) Write down the argument to the right of “outputdir” field:  **OUTDIR** = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| … | Type the following into the console and press enter:  ls [OUTDIR]  where OUTDIR is the value you wrote down in the previous step. | There are NINTERVALS \* NDOMAINS files starting with the prefix “max\_for\_pathway\_for\_time”  There are NINTERVALS \* NDOMAINS files starting with prefix “max\_for\_pathway”  there is one file labeled “testlogfile.txt” |  |

| Table 3  **maxDose Acceptance Test Plan Case 2** | | | |
| --- | --- | --- | --- |
| **maxDose Acceptance Testing**  **CACIE-maxDose – AT-2** | | **Date:** | |
| **Tool Runner Log File Location for this test:**  **[PUT LINK TO THE DIRECTORY HERE]** | | **Test Performed By:** | |
| **Testing Directory: [PROVIDE LINK TO TESTING DIRECTORY]** | | | |
| **Test Step** | **Test Instruction** | **Expected Result** | **Test Result  (Pass/Fail)** |
| *This test assumes that you have already successfully completed Acceptance Test 1. If you have not done so, please complete that test now. If any portion of that test failed, do not start this test.*  *For this test, you must transfer the files from the remote server to a location on your own computer so that you can use Microsoft Excel to independently verify the maximum dose calculation.* | | | |
| 1 | *cd to the OUTPUT directory (OUTDIR) of the previous test case and verify that it contains these four csv files for each boundary (U235):*  ***max for pathway for time******ca98 2070-3070***  ***max for pathway for time******ca98 3070-12070***  ***max for pathway******ca98 2070-3070***  ***max for pathway******ca98 3070-12070***  ***max for pathway for time******inner 2070-3070***  ***max for pathway for time******inner 3070-12070***  ***max for pathway******inner 2070-3070***  ***max for pathway******inner 3070-12070***  ***max for pathway for time******outer 2070-3070***  ***max for pathway for time******outer 3070-12070***  ***max for pathway******outer 2070-3070***  ***max for pathway outer 3070-12070*** | The twelve files were created in the OUTPUT DIRECTORY. |  |
| 2 | *While you can perform the remainder of this test on the machine itself, it might be easier if you the copy the files in step 1 to your local machine and use a program like Excel or similar to inspect individual rows.* |  | NA |
| 3 | Spot check at least five rows in each of the ***max for pathway*** output files against the files with similar names located in the VERIFICATION folder. | The route (**pathway**), **row**, **column**, **layer** and **dose** values all match. |  |
| 4 | Spot check at least five rows in each of the ***max for pathway for time*** output files against the files with similar names located in the VERIFICATION folder. | The route (**pathway**), **row**, **column**, **layer** and **dose** values all match. *Multiple rows exist for some years*. |  |

# Acceptance Test Report

To complete the Acceptance Testing use Appendix A. The two test cases are described as follows:

* Acceptance Test 1 is in Table A-1. This test runs the program and verifies that it accepts the required inputs and produces the required outputs.
* Acceptance Test 2 is in Table A-2. This test asks the user to verify that the output is identical to a file where the calculations for the maxDose were computed by hand independently. Showing that the two files are identical proves that the tool is calculating maxDose as expected.

Details of these tests, when they were conducted, by whom, and if they Passed or Failed are in each table of Appendix A.

# User Guide

This tool is intended to be invoked immediately after ca-dosecalc. We recommend using a shell script to invoke the tool as follows:

# script to test the max dose  
echo "Testing max dose"  
prodTools='/home/ca/CA-CIE-Tools'

controlFile='inputs/controlInput.json’  
python3 $prodTools/pylib/runner/runner.py "python3" "$prodTools/pylib/camaxd ose/maxDose.py $controlFile" --logfile "output/testlogfile.txt"

An example control file is reproduced below. Note that three domains and two time intervals are defined:

{

"copc":"U235",

"dosepath":"inputs/U235.csv",

"comment":"This comment is ignored by this tool",

"domains":[

{"name":"inner",

"fpath":"inputs/Inner\_Area/P2R\_Cells\_On\_and\_Outside\_Outer\_Area\_Bou ndary.csv"},

{"name":"outer",

"fpath":"inputs/Outer\_Area/P2R\_Cells\_On\_and\_Outside\_Inner\_Area\_Bou ndary.csv"},

{

"name":"ca98",

"fpath":"inputs/CA98/P2R\_Cells\_On\_CA98\_Boundary.csv"

}

],

"outputdir":"output",

"dateranges":[

{"start\_year":2070, "end\_year":3070},

{"start\_year":3070, "end\_year":12070}

]

}

# Tool Versions

This section details changes incorporated into each version of the **maxDose** tool.

* 1.0 – Tool was developed.

# Appendix A

**Completed Acceptance Test Cases**

**Tool Runner Log**

INFO--08/06/2020 11:02:35 PM--Starting CA-CIE Tool Runner. Logging to "output/testlogfile.txt"

INFO--08/06/2020 11:02:35 PM--Code Version: 63140375918466bd0d783e985d6fe8e0eda338f4 v4.0: /home/ca/CA-CIE-Tools/pylib/runner/runner.py<--1bcfd6779e9cbdb82673405873a8e5e81514ae27

INFO--08/06/2020 11:02:35 PM--Code Version: 42a66e69e6242a249cb59a8963bb4fe9c68a7ba9 Local repo SHA-1 has does not correspond to a remote repo release version: /home/ca/dose/test-maxDose/CA-CIE-Tools/pylib/camaxdose/maxDose.py<--94e001bfe1b518dd48d20d81d68cf117d5c887a8

INFO--08/06/2020 11:02:35 PM--QA Status: QUALIFIED : /home/ca/CA-CIE-Tools/pylib/runner/runner.py

INFO--08/06/2020 11:02:35 PM--QA Status: TEST : /home/ca/dose/test-maxDose/CA-CIE-Tools/pylib/camaxdose/maxDose.py

INFO--08/06/2020 11:02:35 PM--Invoking Command:"python3" with Arguments:"/home/ca/dose/test-maxDose/CA-CIE-Tools/pylib/camaxdose/maxDose.py inputs/testControlInput.json"

INFO--08/06/2020 11:02:35 PM--Username:ca Computer:twotbbase Platform:Linux 4.15.0-111-generic #112-Ubuntu SMP Thu Jul 9 20:32:34 UTC 2020

| Table A1  **maxDose Acceptance Test Plan Case 1** | | | |
| --- | --- | --- | --- |
| **maxDose Acceptance Testing**  **CACIE-maxDose – AT-1** | | **Date: 8/6/2020** | |
| **Tool Runner Log File Location for this test:**  **/home/ca/dose/test-maxDose/output/testlogfile.txt** | | **Test Performed By: Neira Mondragon** | |
| **Testing Directory: /home/ca/dose/test-maxDose** | | | |
| **Test Step** | **Test Instruction** | **Expected Result** | **Test Result  (Pass/Fail)** |
| You will need to log in to the server to perform this test. Please request the server location and login credentials from the lead developer. | | | |
| 1 | cd into the test directory by typing into the console and press enter:  cd ~/dose/test-maxDose |  | NA |
| 2 | Run the test script by typing into console and press enter:  ./testMaxDoseAT1.sh | The script begins to run and the console notifies you that Acceptance test 1 has started and that the output is logged to output/testlogfile.txt | Pass |
| 3 | Obtain information from the input control file and enter it into the fields in the box on the right:  Type the following into the console and press enter  vim -R inputs/testControlInput.json | 1) Number of unique objects in the “domain” field (each object has a “name” and “fpath” attribute):  NDOMAINS = \_\_\_\_\_\_3\_\_\_\_\_\_  2) Number of unique objects in the “dateranges” field (each object has a “start\_year” and “end\_year” attributes):  **NINTERVALS** = \_\_\_\_2\_\_\_\_  3) Write down the argument to the right of “outputdir” field:  **OUTDIR** = \_\_\_\_\_\_\_output\_\_\_\_\_\_\_\_ | Pass |
| … | Type the following into the console and press enter:  ls [OUTDIR]  where OUTDIR is the value you wrote down in the previous step. | There are NINTERVALS \* NDOMAINS files starting with the prefix “max\_for\_pathway\_for\_time”  There are NINTERVALS \* NDOMAINS files starting with prefix “max\_for\_pathway”  there is one file labeled “testlogfile.txt” | Pass |

| Table A2  **maxDose Acceptance Test Plan Case 2** | | | |
| --- | --- | --- | --- |
| **maxDose Acceptance Testing**  **CACIE-maxDose – AT-2** | | **Date: 8/6/2020** | |
| **Tool Runner Log File Location for this test:**  **N/A** | | **Test Performed By: Neira Mondragon** | |
| **Testing Directory: /home/ca/dose/test-maxDose/** | | | |
| **Test Step** | **Test Instruction** | **Expected Result** | **Test Result  (Pass/Fail)** |
| *This test assumes that you have already successfully completed Acceptance Test 1. If you have not done so, please complete that test now. If any portion of that test failed, do not start this test.*  *For this test, you must transfer the files from the remote server to a location on your own computer so that you can use Microsoft Excel to independently verify the maximum dose calculation.* | | | |
| 1 | *cd to the OUTPUT directory (OUTDIR) of the previous test case and verify that it contains these four csv files for each boundary (U235):*  ***max for pathway for time******ca98 2070-3070***  ***max for pathway for time******ca98 3070-12070***  ***max for pathway******ca98 2070-3070***  ***max for pathway******ca98 3070-12070***  ***max for pathway for time******inner 2070-3070***  ***max for pathway for time******inner 3070-12070***  ***max for pathway******inner 2070-3070***  ***max for pathway******inner 3070-12070***  ***max for pathway for time******outer 2070-3070***  ***max for pathway for time******outer 3070-12070***  ***max for pathway******outer 2070-3070***  ***max for pathway outer 3070-12070*** | The twelve files were created in the OUTPUT DIRECTORY. | Pass |
| 2 | *While you can perform the remainder of this test on the machine itself, it might be easier if you the copy the files in step 1 to your local machine and use a program like Excel or similar to inspect individual rows.* |  | NA |
| 3 | Spot check at least five rows in each of the ***max for pathway*** output files against the files with similar names located in the VERIFICATION folder. | The route (**pathway**), **row**, **column**, **layer** and **dose** values all match. | Pass |
| 4 | Spot check at least five rows in each of the ***max for pathway for time*** output files against the files with similar names located in the VERIFICATION folder. | The route (**pathway**), **row**, **column**, **layer** and **dose** values all match. *Multiple rows exist for some years*. | Pass |

# Appendix B

**Completed Installation Test**

Installation test done as part of acceptance test case.

**Appendix C**

**QA Checklist**

