**CACIE Tool** – **maxDoseDB**

**Version** **1.0**

**QA**: **QA**

# Description and Purpose

This tool is a reimplementation of maxDose that allows for very large input files.

This tool calculates the maximum total dose and the pathway- and COPC-specific contributions to the maximum total doseaover user-specified time intervals and spatial extents. When invoked, maxDoseDB 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 total dose is calculated for each time step, time interval, spatial domain.

# Functional Requirements

The following are the functional requirements (FR) of maxDoseDB

* 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-sumdoseDB 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 total dose results for each time interval and domain specified in the control file.  *Requirements calculating the maximum total dose results are described below.*
  + FR-11: the maximum total dose timeseries results for each domain and time interval specified in the control file
  + FR-12: information relating to where and when the maximum total dose occurred: the year, row, column, and layer.
  + FR-13: for each maximum total dose, information relating to the pathway- and COPC-specific contributions to the maximum total dose
* The tool must compute the maximum total dose by aggregating:
  + FR-14: For each time interval and domain, compute the maximum total dose and the pathway- and COPC-specific contributions over the domain for each year in the interval.
  + FR-15: For each time interval and domain, compute the maximum total dose and the pathway- and COPC-specific contributions.

# Software Requirements Specifications

This tool requires:

* python 3.6+
* PostgreSQL 9+

# 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 maxDoseDB 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":"total",

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

"domains":[

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

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

{"name":"ca98", "fpath":"ca98DomainFile.csv"}

],

“columns”:[

{“name”:”elapsed\_tm”, “type”:”INTEGER”},

{“name”:”model\_date”, “type”:”VARCHAR(100)”},

{“name”:”pathway”, “type”:”VARCHAR(200)”},

{“name”:”cell\_row”, “type”:”INTEGER”},

{“name”:”cell\_column”, “type”:”INTEGER”},

{“name”:”cell\_layer”, “type”:”INTEGER”},

{“name”:”dose”, “type”:”DOUBLE PRECISION”}

],

"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.

* dosepath
  + path to the CSV file containing the dose data produced by the **ca-sumdoseDB** tool.
* columns
  + This is a list of column headers in the target dose file. There must be exactly one entry for every column in the file. Each entry must be an object with the following fields and the order must be identical to the order in the target dose file:
    - name
      * This attribute must match a column name in the dose file.
    - Type
      * must be a valid Postgres primitive type (INTEGER, DOUBLE PRECISION, VARCHAR, etc).
* 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 total 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 total dose and the corresponding pathway- and COPC-specific contributions. 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** or **ca-sumdose**. 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. You tell maxDoseDB how to pare the columns by defining matching correspond to:

|  |  |  |
| --- | --- | --- |
| **Column name** | **Type** | **Description** |
| elapsed\_tm | Integer | The number of days since the start of the model corresponding to this row’s data |
| 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. |
| Soil | Text | The soil type of the grid cell (row/col) |
| Pathway | Text | The exposure pathway |
| cell\_row | Integer | grid index (1-based) |
| cell\_column | Integer | grid index (1-based) |
| cell\_layer | Integer | grid index (1-based) |
| concentration | Float | The concentration, in units of the MT3D UCN file, multiplied by the conversion parameter. |
| dose\_factor | Float | The dose factor for the pathway/soil type  Units are the same as those provided in dosefactsFile |
| Dose | Float | The calculated dose. Equivalent to column 8 \* column 9. |

Output Files:

There are two output files for this tool. The form of the output mirrors the dose input file except the reported dose is the maximum total dose and the corresponding pathway- and COPC-specific contributions for the given spatial domain and time interval.

* **Maximum total dose by pathway as a timeseries**. Files of this type are titled max\_dose\_timeseries.csv. Each row in this file type reports the maximum over the spatial domain DOMAIN and the time interval defined by the START\_YEAR and END\_YEAR for the specified year and pathway. The remaining columns denote the properties of where and when the maximum total dose occurred.
* **Maximum total dose by pathway.** Files of this type are titled dose.csv. Each row reports the maximum total dose for the given pathway over the domain and time interval. The remaining columns denote the time and location of where the maximum total 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/maxDoseDB.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 maxDoseDB tool is presented in Table 1.

| Table  Requirements Traceability Matrix | | | |
| --- | --- | --- | --- |
| **Functional Requirement ID** | **Requirement Description** | **Test Case** | **Pass/Fail** |
| QA Level | Installation Test Case | CACIE-maxDoseDB-IT-1 |  |
| FR-1 | Process a user-specified input control file. | CACIE- maxDoseDB-AT-1 |  |
| FR-2 | Consume the output of the ca-dosecalc tool as input | CACIE- maxDoseDB-AT-1 |  |
| FR-3 | Consume one or more [spatial] domain definition files | CACIE- maxDoseDB-AT-1 |  |
| FR-4 | Specify a directory where the output is stored | CACIE- maxDoseDB-AT-1 |  |
| FR-5 | Define multiple time intervals | CACIE- maxDoseDB-AT-1 |  |
| FR-6 | Define multiple domains | CACIE- maxDoseDB-AT-1 |  |
| FR-7 | Define a domain by indicating a name and a path to the domain input file | CACIE- maxDoseDB-AT-1 |  |
| FR-8 | Define a time interval by specifying a start year and end year in the control file | CACIE- maxDoseDB-AT-1 |  |
| FR-9 | All results in CSV format | CACIE- maxDoseDB-AT-1 |  |
| FR-10 | The maximum total dose timeseries results for each time interval and domain specified in the control file | CACIE- maxDoseDB-AT-2 |  |
| FR-11 | The maximum total dose results for each time interval and domain specified in the control file | CACIE- maxDoseDB-AT-2 |  |
| FR-12 | Information relating to where and when the maximum total dose occurred: the year, row, column, and layer | CACIE- maxDoseDB-AT-2 |  |
| FR-13 | Information relating to the pathway- and COPC-specific contributions corresponding to the maximum total dose | CACIE- maxDoseDB-AT-2 |  |
| FR-14 | For each time interval and domain, compute the maximum total dose and the pathway- and COPC-specific contributions to the maximum total dose over the domain for each year in the interval | CACIE- maxDoseDB-AT-2 |  |
| FR-15 | For each time interval and domain, compute the maximum total dose and the pathway- and COPC-specific contributions to the maximum total dose over the domain | CACIE- maxDoseDB-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  **maxDoseDB Acceptance Test Plan Case 1** | | | |
| --- | --- | --- | --- |
| **maxDoseDB Acceptance Testing**  **CACIE-maxDoseDB – 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-maxDoseDB** | | | |
| **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-maxDoseDB |  | 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) The “domain” fields (each object has a “name” and “fpath” attribute):  **DOMAINS** = \_\_\_\_\_\_\_\_\_\_\_\_  2) The “dateranges” fields (each object has a “start\_year” and “end\_year” attributes):  **INTERVALS** = \_\_\_\_\_\_\_\_  3) Write down the argument to the right of “outputdir” field:  **OUTDIR** = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| 4 | Type the following into the console and press enter:  ls [OUTDIR]  where OUTDIR is the value you wrote down in the previous step. | There is a file “max\_dose.csv”  There is a file  “max\_dose\_timeseries.csv”  there is one file labeled “testlogfile.txt” |  |
| *Copy the files in step 5 to your local machine and use a program like Excel or similar to inspect individual rows.* | | | |
| 5 | In **max\_dose.csv**, verify that for each domain listed in Step 3, there is a time interval corresponding to the intervals listed in Step 3 |  |  |
| 6 | In **max\_dose\_timeseries.csv**, verify that for each domain listed in Step 3, there is a time interval corresponding to the intervals listed in Step 3 |  |  |

| Table 3  **maxDoseDB Acceptance Test Plan Case 2** | | | |
| --- | --- | --- | --- |
| **maxDoseDB Acceptance Testing**  **CACIE-maxDoseDB – 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 csv files:*  ***max\_dose.csv***  ***max\_dose\_timeseries.csv*** | The files were created in the OUTPUT DIRECTORY. |  |
| *Copy the files in step 1 to your local machine and use a program like Excel or similar to inspect individual rows.* | | | |
| 2 | In **max\_dose.csv**, verify the following: | | |
| 2.1 | In **max\_dose.csv,** filter on domain = inner and start\_year = 2070. | Record:  Domain =  start\_year =  cell\_layer=  cell\_row=  cell\_column=  dose associated with Total pathway = | |
| 2.2 | In **totalDose\_check.xlsx**, filter model dates as follows:  2070 ≤ model\_date ≤ 3070 | | |
| 2.3 | In **totalDose\_check.xlsx**, filter on dose = “MAXIMUM VALUE LISTED”  Verify cell\_layer, cell\_row, cell\_column and dose values are equal to the values recorded in step 2.1 | Record value in  cell\_layer=  cell\_row=  cell\_column=  dose = |  |
| 2.4 | For a selected model date in **max\_dose.csv**, verify that the pathway-specific contributions for a specific COPC sum to the total dose for that COPC. | Record selected model date and COPC:  Record pathway-specific dose contributions for the selected COPC  Record the Total value (pathway)for the selected COPC  Verify ∑Pathway dose values = Total dose value |  |
| 3.3 | For a selected model year in **max\_dose.csv**, verify that the COPC-specific contributions for a specific pathway sum to the dose value for that pathway | Record selected model date and pathway:  Record COPC-specific dose contributions for the selected pathway  Record the dose value for the selected pathway  Verify ∑COPC dose values = dose value |  |
| 3 | In **max\_dose\_timeseries.csv**, verify the following: | | |
| 3.1 | Verify the maximum total dose reported for a model date is consistent with the maximum total dose for the same model date in the input file **totalDose\_check.xlsx** as follows:  In **totalDose\_check.xlsx**, filter on domain = inner and model\_date = 1/3/2070 and verify there are more than one cell layer-row-column combination. | Record the cells remaining after applying fitler to verify there is more than one cell in selected model\_date | |
| 3.2 | In **totalDose\_check.xlsx**, filter on  pathway = Total  dose= “MAXIMUM VALUE Listed” | Record values in  cell\_layer=  cell\_row=  cell\_column=  dose column= | |
| 3.3 | In **max\_dose\_timeseries.csv**, filter on:  domain = inner  pathway = Total  model\_date = 1/3/2070  Verify cell\_layer, cell\_row, cell\_column and dose values are equal to the values recorded in step 3.2 | Record values in  cell\_layer=  cell\_row=  cell\_column=  dose = |  |
| 3.4 | For a selected model date in **max\_dose\_timeseries.csv**, verify that the pathway-specific contributions for a specific COPC sum to the total dose for that COPC. | Record selected model date and COPC:  Record pathway-specific dose contributions for the selected COPC  Record the Total value (pathway)for the selected COPC  Verify ∑Pathway dose values = Total dose value |  |
| 3.5 | For a selected model year in **max\_dose\_timeseries.csv**, verify that the COPC-specific contributions for a specific pathway sum to the dose value for that pathway | Record selected model date and pathway:  Record COPC-specific dose contributions for the selected pathway  Record the dose value for the selected pathway  Verify ∑COPC dose values = dose value |  |

# 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 maxDoseDB were computed by hand independently. Showing that the two files are identical proves that the tool is calculating maxDoseDB 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 or ca-sumdose. 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/maxDoseDB.py $controlFile" --logfile "output/testlogfile.txt"

An example control file is reproduced below. It is important that all of the columns in the target input file are defined in the “columns” section and are mapped to the appropriate type. We recommend using DOUBLE PRECISION for dose. Note that three domains and two time intervals are defined:

{

"copc":"total",

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

"domains":[

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

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

{"name":"ca98", "fpath":"ca98DomainFile.csv"}

],

“columns”:[

{“name”:”elapsed\_tm”, “type”:”INTEGER”},

{“name”:”model\_date”, “type”:”VARCHAR(100)”},

{“name”:”pathway”, “type”:”VARCHAR(200)”},

{“name”:”cell\_row”, “type”:”INTEGER”},

{“name”:”cell\_column”, “type”:”INTEGER”},

{“name”:”cell\_layer”, “type”:”INTEGER”},

{“name”:”dose”, “type”:”DOUBLE PRECISION”}

],

"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 **maxDoseDB** tool.

* 1.0 – Tool was developed.

# Appendix

**Completed Acceptance Test Cases**

**Tool Runner Log**