

**When Spreadsheets Become Software - Quality Control Challenges and Approaches –
13360**

Stefanie A. Fountain*, Emmie G. Chen*, Elizabeth E. Wyatt**, John F. Beech*, Tanya B. Quinn***, Robert W. Seifert****, Richard R. Bonczek*****

* Geosyntec Consultants, Inc., 1255 Roberts Boulevard NW, Suite 200, Kennesaw, GA 30144, sfountain@geosyntec.com

** LATA Environmental Services of Kentucky, LLC, 761 Veterans Ave, Kevil, KY 42053

*** Geosyntec Consultants, Inc., 2002 Summit Boulevard NE, Suite 885, Atlanta, GA 30319

**** Portsmouth/Paducah Project Office, United States Department of Energy, 5600 Hobbs Rd, Kevil, KY 42053

***** Portsmouth/Paducah Project Office, United States Department of Energy, 1017 Majestic Drive, Lexington, KY 40513

ABSTRACT

As part of a preliminary waste acceptance criteria (PWAC) development, several commercial models were employed, including the Hydrologic Evaluation of Landfill Performance model (HELP) [1], the Disposal Unit Source Term – Multiple Species model (DUSTMS) [2], and the Analytical Transient One, Two, and Three-Dimensional model (AT123D) [3]. The results of these models were post-processed in MS Excel spreadsheets to convert the model results to alternate units, compare the groundwater concentrations to the groundwater concentration thresholds, and then to adjust the waste contaminant masses (based on average concentration over the waste volume) as needed in an attempt to achieve groundwater concentrations at the limiting point of assessment that would meet the compliance concentrations while maximizing the potential use of the landfill (i.e., maximizing the volume of projected waste being generated that could be placed in the landfill). During the course of the PWAC calculation development, one of the Microsoft (MS) Excel spreadsheets used to post-process the results of the commercial model packages grew to include more than 575,000 formulas across 18 worksheets. This spreadsheet was used to assess six base scenarios as well as nine uncertainty/sensitivity scenarios. The complexity of the spreadsheet resulted in the need for a rigorous quality control (QC) procedure to verify data entry and confirm the accuracy of formulas.

INTRODUCTION

As part of PWAC development, several commercial models were employed, including HELP, DUSTMS, and AT123D. These models were used to evaluate i) theoretical infiltration through the cover system of the landfill; ii) the leaching of contaminants from the waste as a result of the infiltration; iii) the migration of contaminants through the bottom liner system of the landfill; and iv) the fate and transport of the contaminants through the aquifer to a point of assessment.

The modeling methodology for the PWAC development process is depicted in Fig. 1. Climate data (including precipitation) estimated from regional databases was used as the input into HELP. The output from HELP was an estimate of the infiltration through the landfill cover

system into the waste. DUSTMS was then used to estimate the mass of contaminants that would leach from the landfill waste under a given set of assumptions, as well as to model transport of chemicals through the bottom liner system and vadose zone to the aquifer. At this point, the results from DUSTMS were entered into AT123D to evaluate fate and transport through the aquifer to a point of assessment (i.e., a groundwater extraction well). AT123D accounts for both transport and degradation of chemicals. MODFLOW and MODPATH were used as part of a site-wide groundwater modeling program; values from these models are used as inputs to AT123D. (As MODFLOW and MODPATH are employed and evaluated as part of a separate project, they will not be further discussed in this paper). Java Archive script and MS Excel macros were used to translate output files from DUSTMS to AT123D and from AT123D to MS Excel for post-processing, respectively.

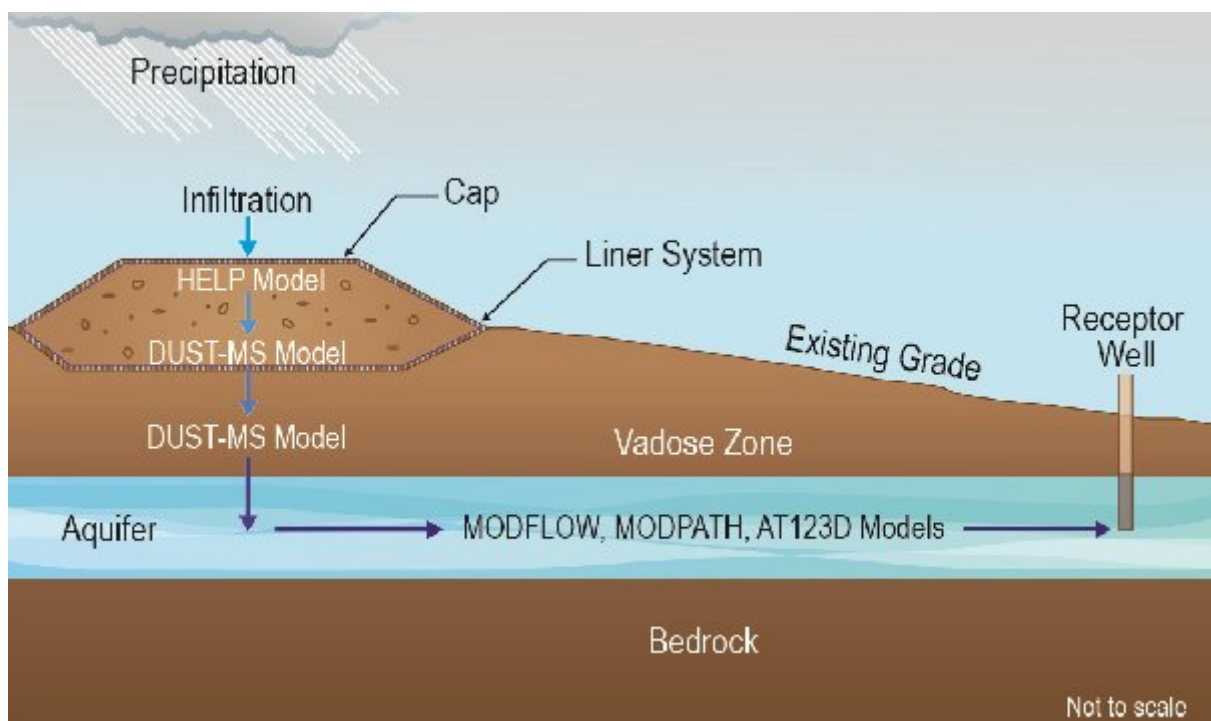


Fig. 1. PWAC Development Conceptual Model

The results of the models were post-processed in MS Excel spreadsheets to convert the model results to alternate units, compare the groundwater concentrations to the groundwater concentration thresholds, and then to adjust the waste contaminant masses (based on average concentrations over the waste volume) as needed in an attempt to achieve groundwater concentrations at the limiting point of assessment that would meet the compliance concentrations while maximizing the potential use of the landfill (i.e., maximizing the volume of projected waste being generated that could be placed in the landfill). MS Excel spreadsheets also were used to compare the calculated PWAC concentrations to the waste projections and provide an estimate of the volume of the projected waste that would not meet the PWAC.

The modeling process is iterative:

- initially, the waste volume was assumed to have 1 mg/kg of each contaminant (106 chemicals were considered);
- the models were run and the concentrations in groundwater at the receptor locations were compared to the project risk criteria;
- the contaminant concentrations in the waste volume were then increased or decreased until the groundwater concentrations at the receptor location were equal to the risk criteria; and
- finally, the projected volume of waste that would not meet the PWAC was determined.

DESCRIPTION OF QUALITY CONTROL PROCEDURE

HELP, DUSTMS, and AT123D Model Verification

Software quality control (QC) was performed using company standard procedures for software verification as well as client QC procedures. In general, these procedures require that the individuals using the software be trained in using the software, that validation and verification (V&V) is performed and documented, that the individuals performing the V&V are qualified to do so, and that the software performs as expected and yields the anticipated results. Hand calculations, results from other analogous software packages, standard problems with known solutions, empirical data, or comparisons to published results may be used for software V&V. For the purposes of this project, V&V of HELP, DUSTMS, and AT123D was performed using comparisons of published results provided by the software developers.

HELP verification was performed using the test case files provided with HELP version 3.07. The files were executed using HELP version 3.07; no modifications to the input files were necessary to perform the verification. The verification model simulation output was saved and compared to the provided output file. The output files had identical results to the provided verification output files.

DUSTMS verification was performed using the provided input files supplied with the installation package (version 3.7). The DUSTMS Windows graphical user interface (DUSTWIN) was used to perform the verification model runs as well as the project simulations. Extensive modifications to the provided files were necessary to convert the input files to the format necessary to interact with DUSTWIN. The chemical concentration/mass output values were verified against the provided output file included with the installation package. Again, the output files had identical results to the provided verification output files.

A Java Archive (“jar” file suffix) file then was used to convert the DUSTMS output files to AT123D input file types.

AT123D was run using SEVIEW licensed from Environmental Software Consultants, Incorporated (ESC) [4]. The installation was modified from the standard AT123D executable file to allow for more load release rates per simulation. After consulting the software developer

ESC, a revised version of AT123D was installed that can handle approximately three times as many varying load release rates per simulation. SEVIEW Version 6.3.12 (with a revised AT123D executable file) was verified using the example files provided in the SEVIEW. Again, the output files had identical results to the provided verification output files.

Spreadsheet QC

Calculation QC packages were prepared for the spreadsheets used for the post-processing of the model results. These included spreadsheets for water velocity calculations, rinse calculations, PWAC calculations, and waste management calculations (i.e., estimation of the volume of waste that would not meet the PWAC). These calculation QC packages consisted of:

- a calculation cover sheet, including the names and signatures of the individuals performing the work and performing the QC;
- a description of the purpose of the calculations;
- a description of the methodology used in the spreadsheet (e.g., the worksheet names and order of calculations);
- a summary of the parameters used for the calculations (e.g., radionuclide half-lives);
- a summary of the equations used and computations performed (e.g., conversion of units, comparison to risk criteria); and
- a summary of representative computation results.

The calculation QC packages for water velocity calculations, rinse calculations, and waste management calculations showed that these spreadsheets were relatively straightforward, with limited numbers of inputs, equations, and calculations. In other words, it would be possible to trace the logic and calculations used in these spreadsheets by reading the calculation QC packages for each.

A review of the calculation QC package for the PWAC calculations indicated a much more complex calculation process was being employed, and an audit was performed to assess whether a calculation QC package process was adequate for the work being performed or if a more involved software V&V should be considered.

DISCUSSION OF PWAC SPREADSHEET AUDIT

Four spreadsheet analysis and audit tools were utilized to conduct an initial assessment of the workbook complexity and to develop an independent list of the potential errors that needed to be considered during QC:

- XLAnalyst (developed by Codematic LTD) - an add-in tool that reviews spreadsheets for frequently used structures that are known to commonly lead to errors and provides a summary of the common metrics to assess the overall complexity of the workbook.
- Arixel Explorer (developed by Arixel Products) - an add-in tool that facilitates a more detailed review of individual formulas beyond the audit functions within MS Excel.
- ExcelSavvy (developed by ExcelSavvy Co.) - a program that provides spreadsheet

auditing tools to better understand models.

- Spreadsheet Professional (developed by Spreadsheet Innovations) - an add-in tool that provides detailed reports of the spreadsheet information to identify for most common errors, general spreadsheet documentation, and other functions.

The audit found that the spreadsheets used for the model post-processing were approximately 10MB in size, with each spreadsheet workbook including 18 worksheets and more than 575,000 formulas comprised of more than 300 unique formulas. The longest formula consisted of more than 260 characters and the most complex formula contained 23 arguments.

The following sources of potential errors were identified by the analysis programs and were considered as items to review in the analysis methodology:

- Mixed Formulas and Values
- Links to External Websites
- Hidden Rows or Columns
- Conditional Formatting
- Array Formulas
- Nested Statements
- Use of SUMIF Statements

Based on the findings of the audit, the PWAC spreadsheet calculations were reclassified as a spreadsheet model. The calculation QC packages previously prepared for the PWAC calculation spreadsheet model, while providing important information and context on the calculations being performed, were considered inadequate to fully QC the complex spreadsheet model.

DISCUSSION OF REVISED QC METHODOLOGY AND RESULTS

Overall Review

Given the magnitude and complexity of the spreadsheet model, an independent QC team of seven (7) individuals with varying backgrounds in quality control, project delivery, environmental engineering, hydrogeology, groundwater modeling, and programming was assembled. The QC team consulted with the project team to assess the critical contaminants that drive conclusions for the project. Discussions with the project team led the QC team to select QC targets consisting of four (4) outputs of the spreadsheet model for four (4) contaminants identified for QC review. The four (4) contaminants were selected as being representative of the main contaminant categories projected to be present in the waste.

The suitability and accuracy of the equations utilized in the spreadsheet were reviewed independently and separately from the QC review of the spreadsheet model. This review process was facilitated through the development of formula and variable mapping that visualize the calculation routine for the QC targets by showing how the dependent variables were sequentially calculated. These maps are called Calculation Summary Maps and an example for

the QC targets is included as Fig. 2. The map shows how dependent variables were sequentially calculated to obtain the PWAC for one of the points of assessment (the edge of waste, or EOW). Development of this map provided the framework of how equations in spreadsheet cells were linked, which in turn provided the basis for how to verify that cell formulas and calculated values were correct. This process was repeated for the other points of assessment (i.e., potential receptor locations) used during the project.

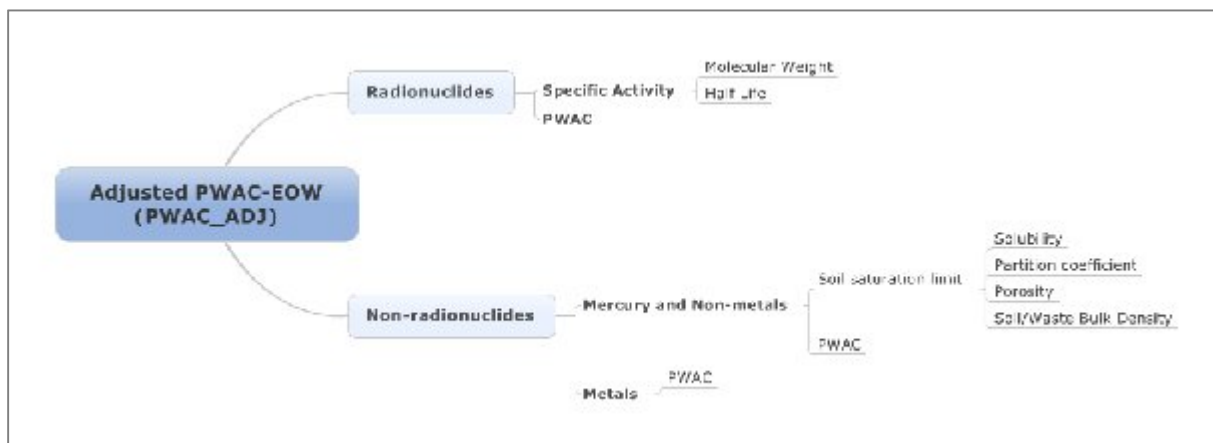


Fig. 2. Example Calculation Summary Map

Worksheets that contained input data for the calculations in the spreadsheet were independently validated by the project team and independent technical review teams. Therefore, for the purposes of this QC review, it was assumed that the input values were obtained from appropriate data sources and appropriate QC checks were previously completed.

Worksheets that served the purpose of performing PWAC calculations were evaluated by the QC team in closer detail with spreadsheet audit tools using visualization to map formula logic. Initial overall checks included the following steps:

- confirm that the logic between the related worksheets was accurate;
- review audit tool visualization outputs for each worksheet to check for anomalies and evaluate the validity of these changes; and
- confirm that external links to the two other spreadsheets referenced within the spreadsheet were valid and being updated properly.

Detailed Spreadsheet Analysis

The process below describes how the QC targets were evaluated in detail for the four (4) selected contaminants.

The QC targets were calculated using a string of formulas embedded within the multiple worksheets in the spreadsheet model. To establish an understanding of how the model parameters were sequentially calculated in the spreadsheet, Calculation Summary Maps (formula

visualizations) were developed showing dependent variables used for the sequence of calculations. The relationships of the input and output variables for each formula were mapped based on the spreadsheet model (see Fig. 2).

The QC team also reviewed the previously prepared calculation QC package. The PWAC calculation formulas in the prior calculation QC package were compared against the corresponding formulas in the spreadsheet model to verify that each formula was correctly input into the spreadsheet model. Inconsistencies between the calculation QC package and the spreadsheet were hand documented by the QC team in the calculation QC. The inconsistencies documented did not result in changes to the spreadsheet model results.

The Calculation Summary Maps were further developed into a Detailed Variable Map showing formula components, input and output variables, and the corresponding worksheets, rows and columns.

For each of the four (4) QC Target contaminants, the formulas and variables were inspected and manually followed through the spreadsheet cell by cell, using the Detailed Variable Maps. An example Detailed Variable Maps is depicted in Fig. 3. These maps were used to manually check that the input and output value of the calculation routine was pointing to the intended location, and the value of each calculation step was documented on the Detailed Variable Map.

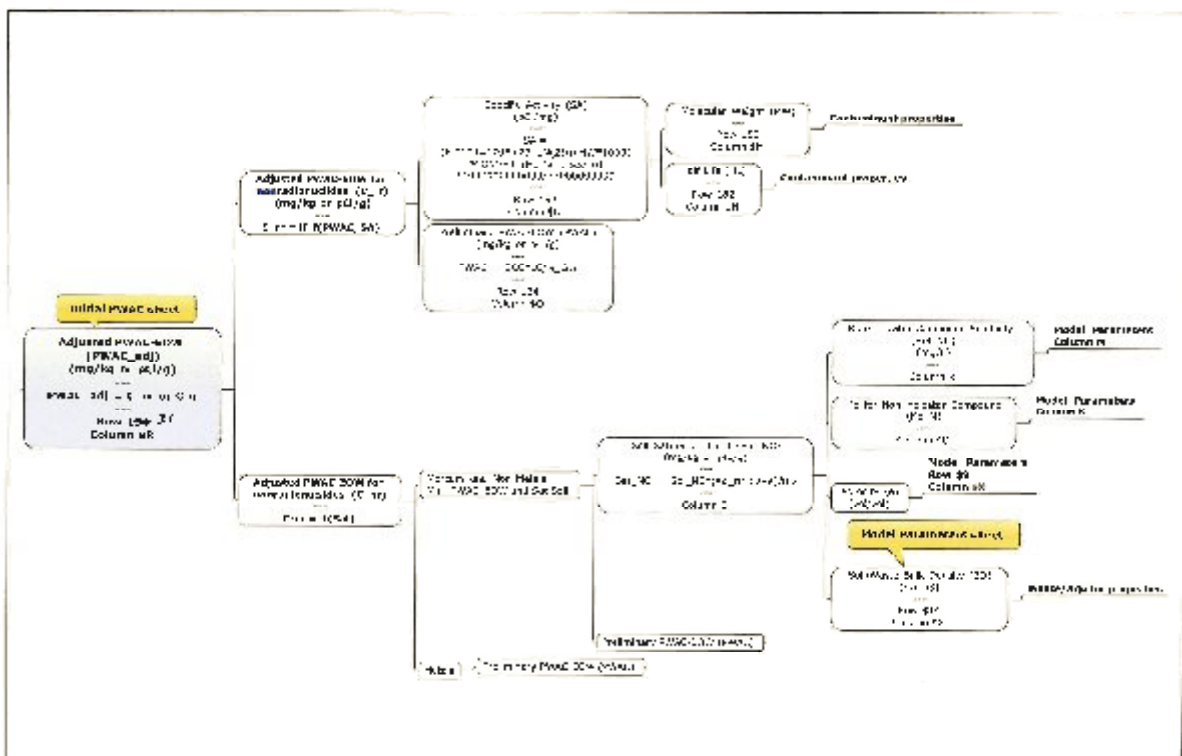


Fig. 3. Example Detailed Variable Map

The variable validation process also included the following checks to the worksheets that were part of the calculation routine for the QC targets:

- confirmed the integrity of units for the calculations;
- confirmed that column and row headers match data being referenced in other worksheets via the lookup function to data arrays;
- confirmed that data array lookup functions returned the intended value; and
- confirmed that the functions were written to include appropriate rows or columns (i.e., if a function was written to pull in groundwater concentration data contained in columns A:D, then the function was written to pull data from columns A:D and not A:C).

QC Results

A few housekeeping items were identified during the spreadsheet review including the following:

- Some column headings did not match the units of the calculations. The QC team was able to confirm that the column headings needed to be revised; however, the formulas and model in the body of the worksheet were accurate.
- Some data sets in the worksheets are not used in formulas. While not a potential error source, excess data not required for the model is information clutter and not generally considered a best practice.

The calculation concepts were verified using the formula mapping and visualization tools. The QC team confirmed that conceptual sequence of calculations related to QC targets and, subsequently, the dependent variables are accurate.

Detailed verifications of the formulas were conducted for the sequence of calculations for the QC targets utilizing visualization mapping for the input variables through the output variables. First, the prior calculation QC package was reviewed in detail and the formulas in the package were reviewed and verified to be input correctly into the corresponding cells within the spreadsheet.

Through an independent process, formulas for the QC targets within the spreadsheet were then cross-referenced with the prior calculation package. During this process, the following issues were identified:

- The landfill concentration and landfill mass calculations used within the spreadsheet for non-radionuclides did not have corresponding explanations in the prior calculation package. An independent review was conducted to confirm that the calculation in the spreadsheet was correct.
- The soil saturation limit calculation used within the spreadsheet for non-radionuclides did not have a corresponding explanation in the prior calculation package. An independent confirmation was conducted to confirm that the calculation in the spreadsheet was correct.

- A unique soil saturation limit calculation was supposed to be applied to a particular contaminant, but was incorrectly applied to a different contaminant. This incorrect formula had no impact but the implementation of the soil saturation limit calculation was communicated to the project team for revision.
- An incorrect syntax was identified within an extended If/Then statement. The same formula was identified as the largest formula in the spreadsheet model by one of the spreadsheet audit tools. The incorrect syntax was reviewed and the result was a non-impacting redundancy within a logic statement.

The QC team reviewed 57 unique formulas during this process. This exceeds the industry best practice (as defined by the General Services Administration) for spreadsheet validation, which suggests that 10% (32 in the case of this spreadsheet) of total unique formulas should be reviewed.

Sixteen (16) Detailed Variable Maps were validated to confirm that the formulas and variables pointed to the intended location and value. Detailed Variable Maps for each QC target were developed and cross-checked. The calculations, variable inputs, and variable outputs were determined to be accurate. The unique variables for the other constituents were manually added to the maps and validated. Calculations, variable inputs, and variable outputs were accurate and no errors were identified.

The QC team reviewed 41 unique variables during this process. This list includes model and site parameters as well as dependent variables that are formulaic based on other variables.

CONCLUSIONS

During the course of a project, calculation packages often evolve as a result of new information, regulatory input, improved technical understanding of the conceptual model, etc. Care should be taken to assess the status of spreadsheets used to perform calculations so as to monitor their potentially increasing complexity during a project. Several commercial tools are available to assist in the spreadsheet audit process and to help users form or apply appropriate QC procedures, thus verifying the conclusions of the calculations performed.

REFERENCES

1. Schroeder, P. R., T. S. Dozier, P. A. Zappi, B. M. McEnroe, J. W. Sjostrom, and R. L. Peyton. 1994. The Hydrologic Evaluation of Landfill Performance (HELP) Model: Engineering Documentation for Version 3. U.S. Environmental Protection Agency Office of Research and Development, Washington, DC, September, EPA/600/R-94/168b.
2. Sullivan, T. M. 2001. Disposal Unit Source Term: Multiple Species (DUSTMS) Data Input Guide. Brookhaven National Laboratory, Upton, NY.
3. Yeh, G.T. 1981. AT123D: Analytical Transient One-, Two-, and Three-Dimensional Simulation of Waste Transport in the Aquifer System. Publication No. 1439, Environmental Sciences Division, Oak Ridge National Laboratory: Oak Ridge, TN, ORNL-5602.

WM2013 Conference, February 24 – 28, 2013, Phoenix, Arizona, USA

4. Schneiker, R. A. 2006. SEVIEW Integrated Contaminant Transport and Fate Modeling System, User's Guide, Version 6.3. Environmental Software Consultants, Inc., Madison, Wisconsin.