Review of BASINS 2.0

prepared for

The Cadmus Group, Inc. & U.S. Environmental Protection Agency

by

William W. Walker, Jr., Ph.D., Environmental Engineer
1127 Lowell Road, Concord, Massachusetts 01742-5522
Tel: 978-369-8061, Fax: 978-369-4230
e-mail: wwwalker@shore.net
http://www.shore.net/~wwwalker

September 14, 1998

Introduction

My review process involved the following steps:

- downloading & reading documentation
- reading course notes provided by Cadmus (apparently for previous version of Basins)
- downloading and installing data files for the tutorial and three other accounting units (two in Massachusetts & one in Michigan)
- exercising each of the Assessment Tools, Utilities, and Reports in each watershed
- exercising the QUAL-2E & NPSM modeling applications
- preparing written comments

Relevant experience or lack thereof includes the following (see web site referenced above for details):

- no previous experience with BASINS or HSPF
- reasonably fluent with ArcView; no experience with Avenue or ArcInfo
- developed P8 Urban Catchment Model for Rhode Island DEM & USEPA
- developed GIS-Based Phosphorus Loading Model for Vermont DEC
- developed FLUX, PROFILE, & BATHTUB, software for evaluating and predicting eutrophication problems in reservoirs
- developed PONDNET, spreadsheet for routing phosphorus through watershed/pond networks.

- assisted EPA in developing early versions of QUAL2E (nutrient & algal kinetics)
- emphasis on eutrophication problems in lakes, reservoirs & wetlands
- experience with BOD & nutrient load allocations in rivers & impoundments, but not with TMDL's per se

Tests were conducted on the following computer system:

- Dell 300 Mhz Pentium Pro
- 64Mb Ram
- 132 Mb Swap File
- Windows 95, Version 4.00.950 b
- 100 mb free disk space (beyond swap file), after installing program & data files for three cases

Although my computer crashed several times in the process of working with BASINS, I was unable to reproduce these episodes and the frequency was probably not different from that typically encountered when working in Arc View. I found that crash risk decreased if I closed and restarted BASINS occasionally, rather than working with it continuously for long periods of time. This a typical remedy for Windows 95 users.

The structure of the review follows the BASINS documentation:

- Installation
- Assessment Tools
- Utilities
- Watershed Characterization Reports
- Stream Water Quality Models
- Nonpoint Source Model
- Data Files
- General

Page, Table & Figure numbers cited below refer to BASINS 2.0 documentation.

References to web links are abbreviated as follows:

XXX/install.htm == http://www.epa.gov/ostwater/BASINS/install.htm

My overall impression of BASINS is a positive one. Please excuse the fact that my comments may be overly severe and critical in some cases. I assume that critical feedback is important and useful for improving the product. My review is somewhat handicapped by the short time allowed for the project.

The short duration has made it difficult to conduct the project and meet other obligations.

Installation

The software and data files were downloaded without problems. Downloading instructions provided at the web site were clear.

After installing BASINS software, a glitch was encountered in installing ArcView Dialog Designer.   The following error message appeared:

unable to run a program needed to proceed with installation.
*restart your system. error 105.*

I rebooted and repeated the entire installation (SETUP) process. The message did not appear a second time.

BASINS 2.0 requires ArcView 3.0a, which I had previously installed.   I recently purchased ArcView 3.1, but have not installed it because I was concerned that it might interfere with my review.   Obviously, compatibility with ArcView 3.1 is an important consideration.   Compatibility with 3.1 (or lack thereof) should be described in the documentation and at the web site.   If it is not compatible, this would pose a problem for users who want to take advantage of new ArcView features.   Maybe it is possible to have 3.0a and 3.1 installed simultaneously on the same computer, though this solution would not be very attractive.

After downloading and working with data from three separate accounting units (USGS 8-digit Hydrologic Unit (HU) codes), it became apparent that some BASINS features are not applicable when the data are cut in this way. For example, the 'Target" tool is not particularly useful because each project is apparently restricted a single HU.   Is there a way of combining data from more than one HU (separately downloaded) into a single ArcView project? Otherwise, there is apparently no way to use all BASINS features until the data files are provided on a state and/or regional basis. Is this correct? If so, it should be stated in the documentation and at the web site.

Massachusetts is spelled wrong at *XXX/STATES/MA*.

I extracted data and built separate projects for each of three hydrologic units. These steps went smoothly (at least I thought so, until I started exercising BASINS features).   Problems were apparently encountered during the projection step of the data extraction process.   After responding 'Yes' when asked whether to project the data, the following dialogue box appeared:
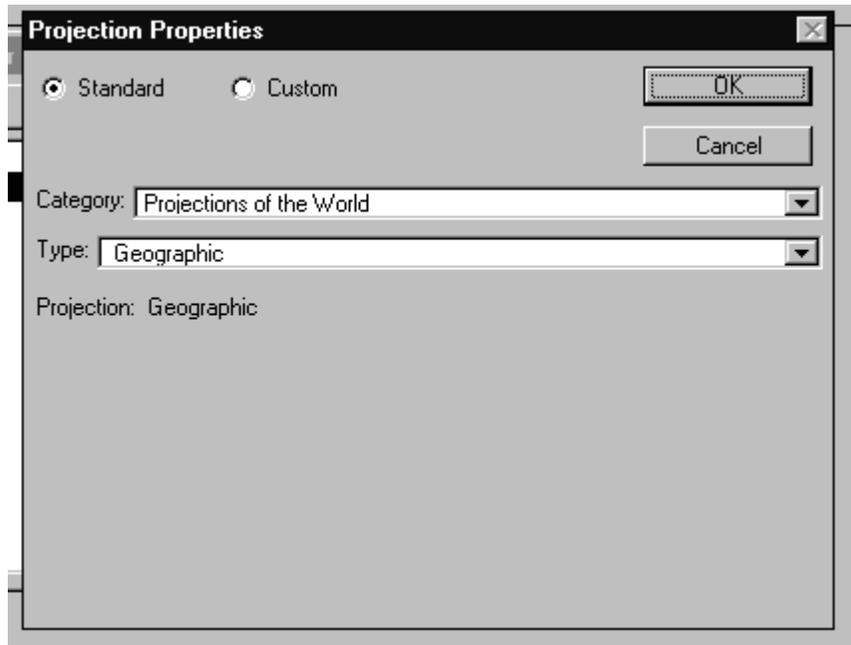
After clicking OK, the extraction process continued, apparently without problems. I working with the project file, however, I found that the data were apparently not projected, as requested. When I selected 'Lookup Projection Parameters' from the BASINS menu, the following error message appeared:
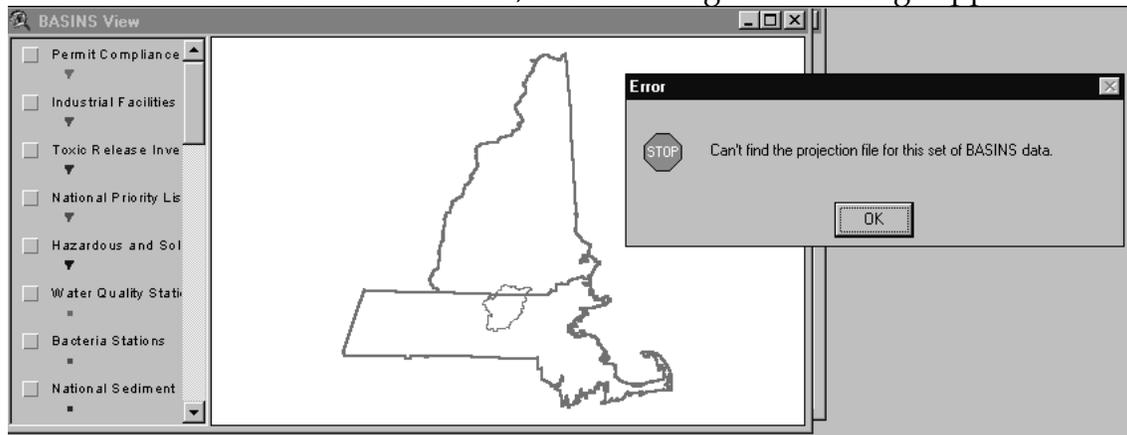
I also found that report utilities involving computation of areas (Landuse Distribution, State Soil Characteristics, Watershed Topographic) would not work (reported areas = 0) with this project. The same problems were found when two other hydrologic units were extracted in a similar fashion.

I found that data projection would occur during the extraction process only if I selected something other than the default Category (projections of the world) and Type (geographic), for example:
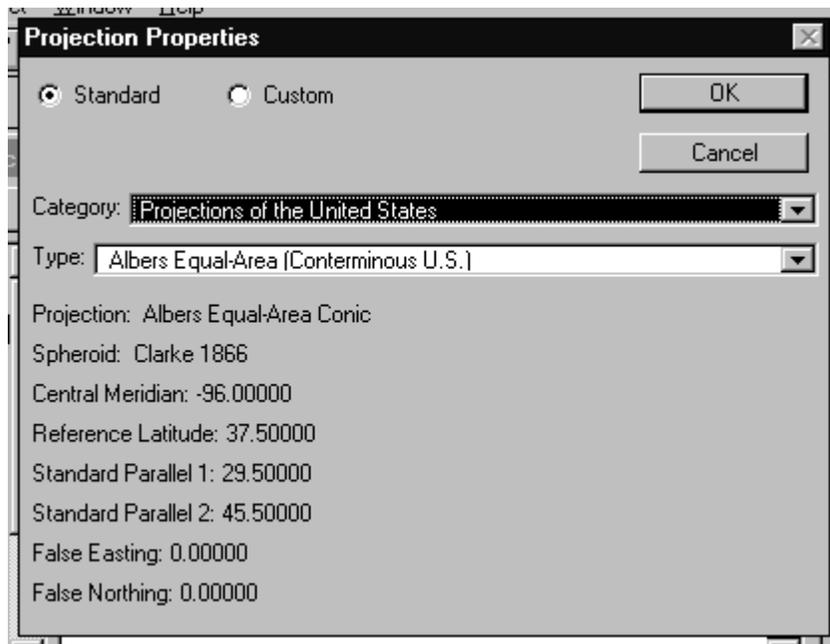
It is not clear whether this is a bug or an intended feature that requires documentation.   In any case, the concept of 'projection' is foreign to me and I would have no basis to make an appropriate choice.   Guidance at the web site: "c. Select a projection, enter appropriate projection parameters and click OK in the Projection Properties dialog box." is not helpful.   What are reasonable choices? It seems that the default choices are poor ones, since they do not project the data.

If an existing data set is re-extracted, the following message box appears:
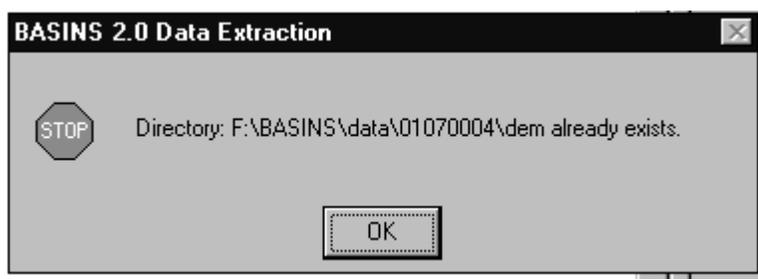
The box indicates "Stop", but after clicking 'OK' (the only choice), the extraction process continues anyway. This is apparently a warning message, but the user is offered no choice but to continue. An optional "Cancel" button seems appropriate here.

When the data were projected using the above options (Projections of the US, Albers Equal-Area), the map image is not oriented in properly:
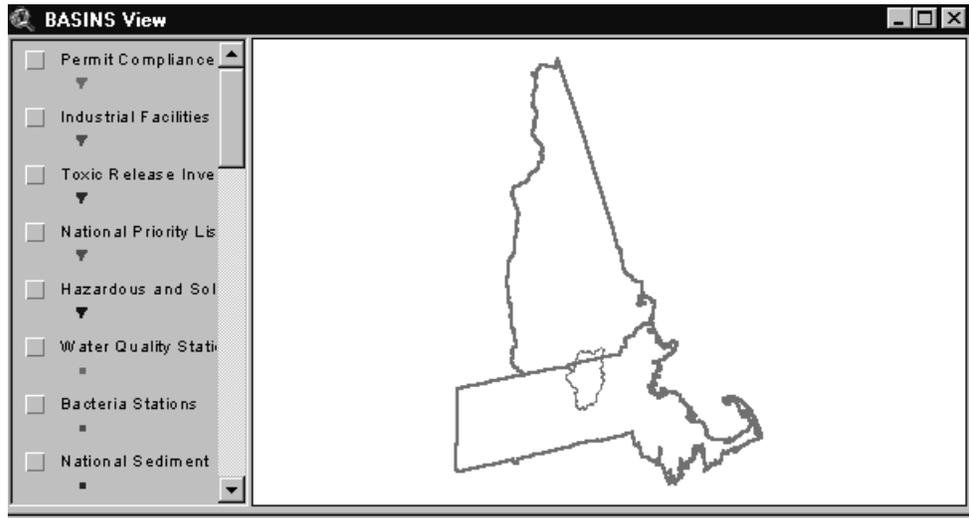
The tutorial case (which had already been extracted and projected before downloading) has a similar orientation problem. This is not acceptable if BASINS is to be used to produce maps and figures for reports or public consumption. Most users will not be sufficiently versed in GIS jargon or software to know how to correct this. Clearly, more help and better default choices are needed for projections.

Oddly enough, when a description of the projection is requested from the ArcView Menu (View/Properties/Projection), the results are "Projections of the World, Geographic", even though the data had been extracted using "Projections of the USA, Albers-Equal Area". When I tried messing with these settings, the project became unusable, probably because I did not know what I was doing. More guidance is needed on projections.

Since projection of the extracted data is apparently required for some BASINS functions, it is not clear why projecting the data is presented as an option during the extraction process. Shouldn't it be done automatically?

Projection and datum compatibilities should be considered when planning to combine BASINS coverages with user-supplied coverages. More guidance is needed in this area. For example, if I have an independent coverage, how would I find out the projection and datum? Would I need ArcInfo in order to change the projection of a given coverage or can this be done in ArcView?

I was able to add RF3 and DEM files to each project using the ArcView 'Add Theme' procedure. This went smoothly. I noticed that the DEM coverage consistently exceeds the HU boundary. While it appears to be intentional, but some discussion of this seems appropriate in the documentation.

I noticed that the RF3 and RF1 coverages were offset slightly in each test case. The RF3 coverage appeared to align better with the water quality station locations than the RF1 coverage.

The RF1 coverage appears to determine potential model applications. The coverage is too coarse (and apparently inaccurate) to be very useful in the watersheds that I downloaded. Modification of modeling functions to work from RF3 coverages should be considered.

Assessment Tools

Some control over the scales used to create histograms in each assessment may be helpful. The interval scales are unequal and not rounded off. This is probably a limitation of ArcView.

The Target application was not particularly useful because it is apparently intended to provide comparisons across hydrologic units (HU's) and the downloaded projects were each restricted to one HU. The meaning of the

It would be helpful if the meaning of the "threshold value" were more clearly defined on the following input screen:
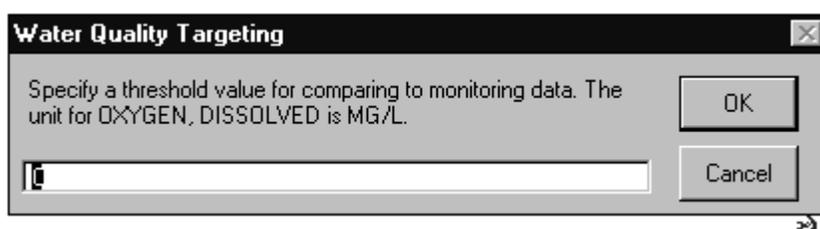
I found that the 'Assess/ Water Quality' tool was most useful for identifying stations with data for a particular parameter and year interval. Selecting the 'Identify Results' tool and pointing to individual stations on the map produced by the procedure produced a table describing station characteristics (see below). The output table contains the average value for the station. It would also be helpful if the table indicated the number of measurements used to compute the average value.
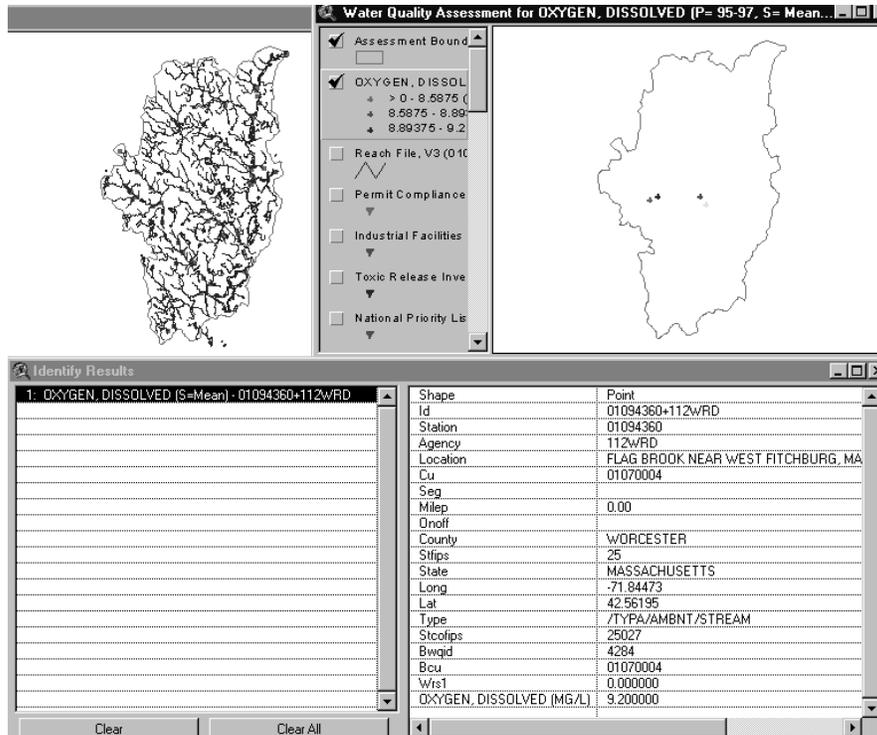
A way of distinguishing between station types (lakes, rivers) would be useful in BASINS assessments. Greater flexibility in selecting date ranges, seasons, depths, etc. would also be desirable, see **Data** discussion.

To aid in identifying data-rich stations suitable for supporting modeling efforts, it would be nice to be able to specify a minimum number of observations required in order for a particular station/parameter combination to appear in output tables and maps produced by the Assess and Data Mining routines.

The Data Mining utility works as advertised, although I was initially stalled by the fact that it does not appear in the program menu and can be accessed only via the Data Mining button.

When initiating Data Mining, I found that it was important not to select particular watershed(s) beforehand. The selection is done by station using the Data Mining tool. If a watershed is selected beforehand, selected stations (Yellow) do not show up in the Data Mining window because the watershed background is also yellow.

I found that selecting specific stations with the "Select Feature" tool in the Data Mining window was a particularly effective way of exploring data coverages. Highlighted rows in the various tables provide the monitoring-station

description, parameter coverage, and date ranges. If I were using this in an actual TMDL project, it would probably be more useful for identifying relevant stations than for accessing and manipulating the actual data. I would probably use this tool to select relevant stations and then submit a separate STORET retrieval request to obtain the raw data for further analysis and modeling outside of BASINS. As discussed below (see Data), BASINS might be configured to facilitate data selection and submission of STORET retrievals.

Although the linkage provided by the Data Mining procedure is useful, there is an information glut on the screen and it is difficult to grasp immediately. The information is there, however, and users will presumably learn how to dig out whatever they are looking for.

Utilities

In running the Watershed Delineation tool, I always received the following message after adding a new watershed:
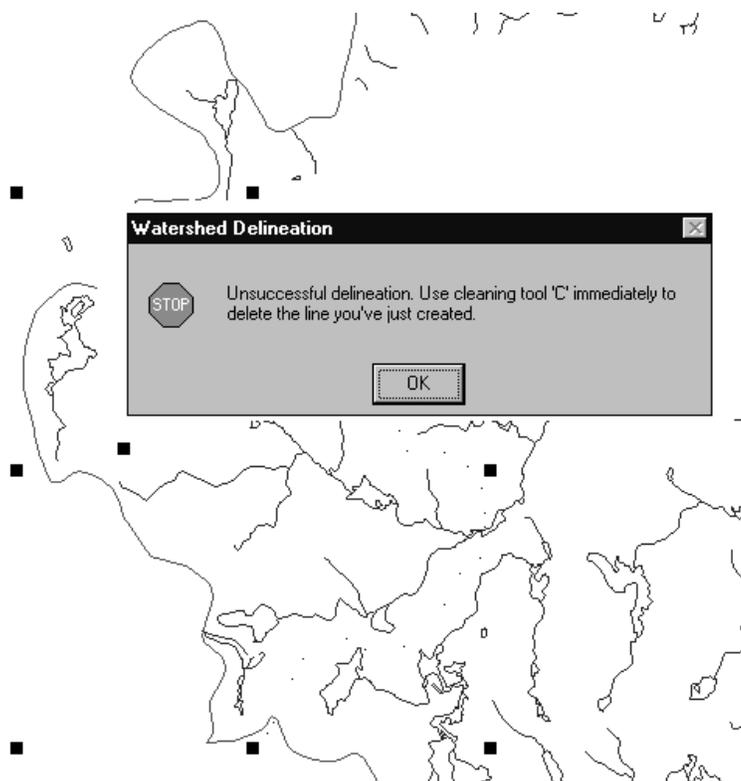
After clicking 'OK', however, the new watershed consistently appeared:

The new subwatershed was defined and recognized in reporting & assessment procedures, even though the error message appeared. I tested the delineation procedure on several other subwatersheds and usually obtained the same error message. The message seems to be correlated with whether or not the delineated subwatershed contains a lake segment. If the delineation tool works only with reach networks that do not contain a lake, this should be indicated in the documentation.

The Watershed Cleaning tool does not work as quite as advertised. After selecting the 'C' button and clicking on the polygon that I wanted to remove, I got the following message:
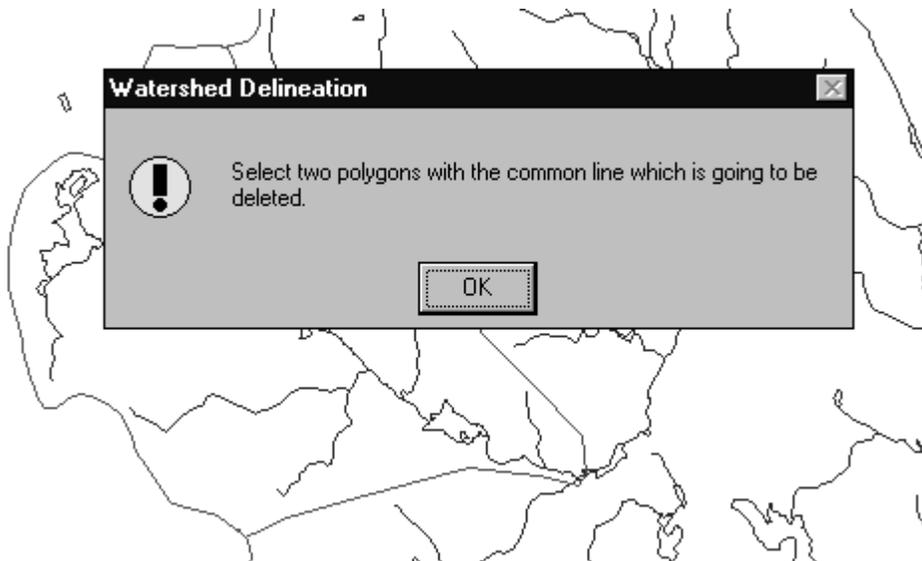
All that was necessary was to draw a box around the polygon that I wanted to remove and respond 'Yes' to the prompt:
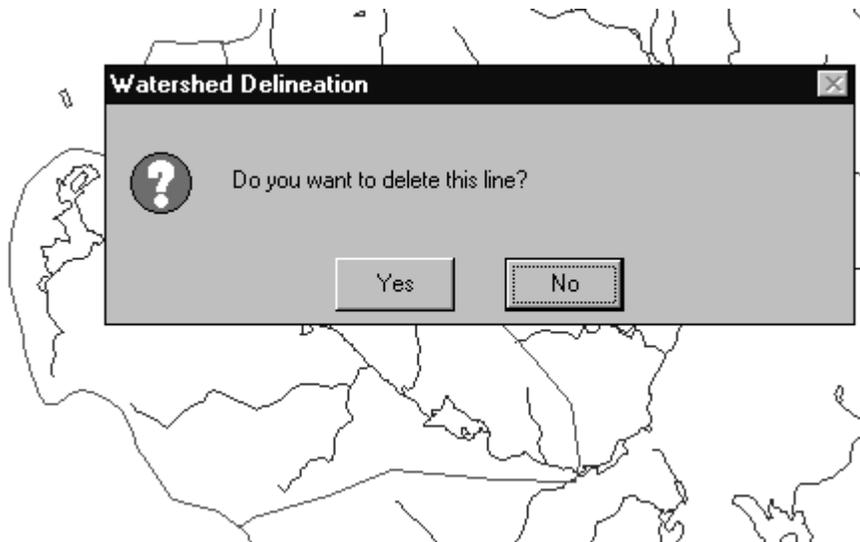
This successfully removed the new segment:

Sometimes, the watershed cleaning tool gets stuck in some kind of loop. After clicking on the cleaning tool and attempting to select a subwatershed for deletion, the following message appears:

```
Watershed Delineation                                        ✕

   (!)      Please click the Clean button first.


                        ┌──────────┐
                        │    OK    │
                        └──────────┘
```
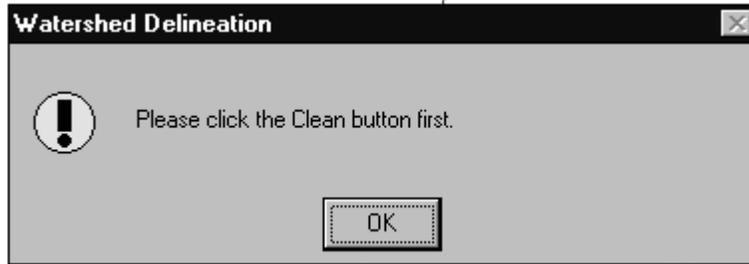
Following instructions does not help. The above message appears repeatedly
after clicking the clean button and attempting to select a subwatershed.   The
entire program crashes shortly after this problem develops.

I tested the land use import function by copying the BASINS supplied landuse
coverage to a separate directory and renaming all the files.   More specific
instructions are needed (rel to Tip on page 7.3-1) on which files are to be copied
and how they are to be renamed. I copied all of the lu files to a new directory
under the current HU and renamed each (e.g., *lu_bostma.** to *lu_bostma2.**). This
is probably wrong. It is not clear what to do with the file called *lulc_fl.src.* This
file contains a line with a file reference that will be incorrect once the other files
are renamed.

The land use import utility appears to be confused/confounded with the land
use reclassification utility, both in the program and in the documentation (pp
7.2-4 to 7.2-6).   The contents and order of user prompts during an import
session is different from that shown in the documentation (pp. 7.2.5-6). It is
not at all clear why the user is asked to re-define land use codes when he is
attempting only to import a new coverage.

The imported coverage came across with a Theme Name 'L_bostma2.shp G
Land Use Group00'.   I have no idea what this means or how it is to be used. I
tried to reclassify one of the codes and got part way through the process when I
encountered error messages and screen prompts that I did not understand. The
documentation is equally confusing.

I gave up on further testing of the land use import/reclassification functions
because it became apparent that it would take a long time to figure them out
and I was running out of time to complete review of other basins features.
These features need serious revisions to the software and/or documentation in
order to make them comprehensible and useful.

In displaying the land use coverages (BASINS-supplied or other), it would be
nice to be able to clip the coverages at the borders of the watersheds being
evaluated, instead of viewing the whole tile.

I tested the 'Reclassify DEM' Utility and found it to work satisfactorily.   It is not clear, however, exactly how this utility would be used in a modeling effort, which parameters one would adjust to achieve what desired effects and why.

I tested the "Append Water Quality Observation" Utility by reading the BASINS supplied observation file into Excel, deleting all records, entering a few new ones, and saving the file in DBF format with a new name. This provided template for the file structure. When I ran the utility, however, I received the following message:
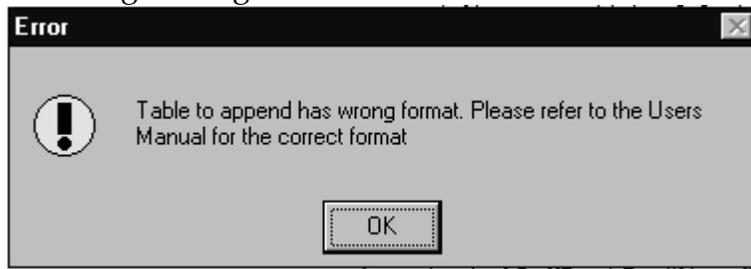
Apparently, DBF files created by Excel are not compatible with the BASINS import utility. I tried this again exporting the file from Excel in both DB III and DB IV formats.   The same results were obtained. I gave up and moved on to testing other features.

Even if I could get the append utility to work, the documentation does not describe the significance or requirements for the 'ID_SAMP' and 'ID' fields. Do these have to be in a particular format or sequence? Do the appended records have to be sorted in any particular way? What happens if the 'ID_SAMP' or 'ID' fields duplicate values in the original file?

Another way of augmenting records would be to load the entire BASINS-supplied data file into EXCEL, cut/paste new records onto the end, and save it in DBF format. Assuming that the format problems described above could be resolved, would this work?   Does this file have to be sorted in any particular way?

I tested the 'Move Station', 'Edit Station' , & 'Delete Station' utilities and found them to work satisfactorily. I found that the 'Edit Station' utility was a quick way of reviewing characteristics of individual stations (even if was not interested in changing them).   This is an alternative to using the ArcView 'Identify' tool.

The format of data files produced by the "Export Water Quality Observation Utility" is fairly clear but not particularly conducive for importing to spreadsheets for further processing.   Perhaps a comma or tab-delimited file

would be more appropriate. It is particularly difficult when dealing with more than one parameter and/or station. The output data set does not contain the remark field, which is in the source data file and could be important in some applications. The list of potential applications for the exported file (p. 7.4-2) is impressive. These applications are indeed critical. Unless BASINS provides a more direct link to such applications, however, I would probably use BASINS only to identify appropriate stations and then submit a separate STORET request to obtain the data in a format that is more comprehensive and easier to work with.

Another more direct way of accessing the Observation file would be to load the entire file into an external spreadsheet or database program and then extract relevant records. Particularly for larger problems, this would probably be easier and more reliable than going through the export utility because it would avoid having to read and parse the ASCII output file.

Reports

I tested the 'Point Source Inventory' and 'Water Quality Summary' reports and found them to work satisfactorily. In generating Water Quality reports, I sometimes found that it was necessary to clear all selected date ranges first and then re-select the desired date ranges in order for the program to properly locate and summarize the water quality data. To assist in identifying trends, it would be nice to be able to plot concentration values (25%, 50%, 75%) vs. date interval at selected stations, but it is not clear whether ArcView could be configured to do this.

It is unfortunate that the land use distribution report tool does not work with imported land use coverages. This limits uses of BASINS to develop input data files for models other then HPSF. For example, tables showing land use (and soil type) distribution by subwatershed would be useful for developing input data sets for applying export-coefficients or other models such as Schuler's "Simple Method" or P8. Such applications are apparently limited to the original land use coverages supplied with BASINS.

I found that the land use report could be applied separately to new subwatersheds defined using the watershed delineation tool. This seems very useful. If I selected two or more user-defined subwatersheds before requesting a land use report, the columns of the report each contained the total area in each category instead of the area within each subwatershed, as illustrated below.

Selected sub-watersheds:

Resulting land use distribution report:

```
Landuse Information - Summary by Major Land Use Category
Table xx. Land use distribution by major land use category.


Sub-Watershed                    01070004001   01070004001   Total Area
=====================================================================  ==========
Land Use Name and Code           Area (acres)  Area (acres)  (acres)
---------------------------------------------------------------------  ----------
Urban or Built-up Land           4204          4204          4204
Agricultural Land                3956          3956          3956
Forest Land                      46316         46316         46316
Water                            1644          1644          1644
WetLand                          478           478           478
Barren Land                      614           614           614
=====================================================================  ==========
Total                            57212         57212         57212
```

It seems that the land use report is only applicable to a single user-defined
watershed (or group of watersheds) at a time. It would be useful to be able to
obtain a breakdown similar to that which is apparently obtained when the
report is requested for multiple hydrologic units. In order for this to work,
there would have to be a way for users to assign a different identifying code to
each user-defined subwatershed. It appears that similar constraints apply to
the soil distribution and watershed topographic reports.

When I attempted to create a soil distribution report (water table depth) for the
above selected sub-watersheds, I obtained the following error message:

This message seems to indicate a problem only with the output layout screen, not with the tables.

I really liked the ability provided by the soil distribution report to generate maps separately for each soil characteristic. I also found it useful to display output layouts from the land use & soil distribution reports side-by-side:
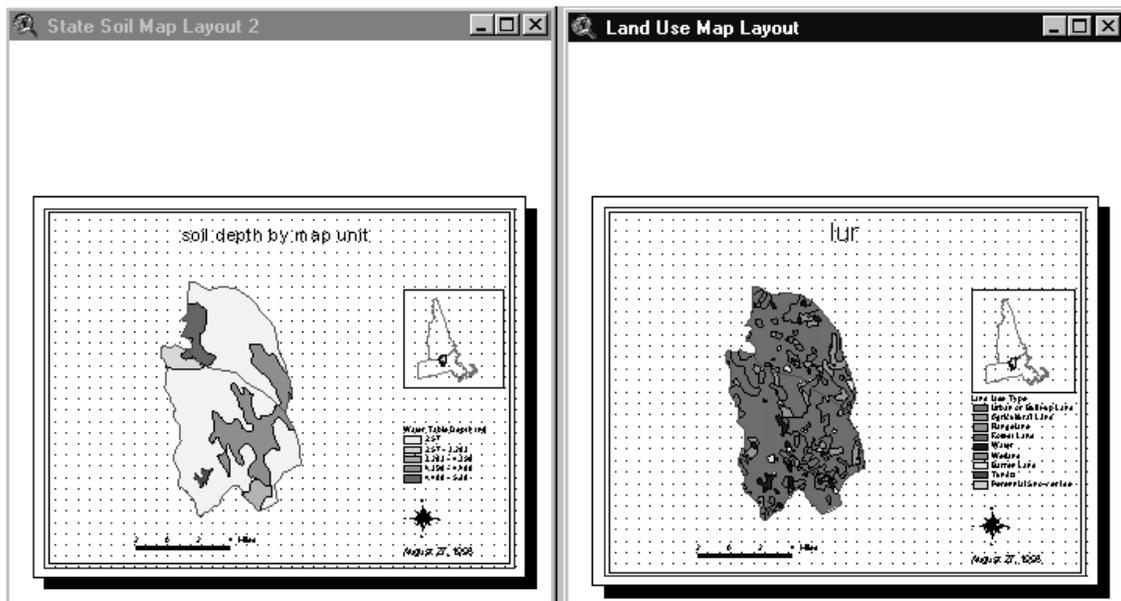
The watershed topographic reports seem to function as described. More guidance is needed for interpreting the output, however. For example, the documentation (p. 8.6-1) states: "This information can be used to quickly evaluate the relative 'Steepness' of the watershed compared to that of other watersheds and correlate it with the results of water quality modeling". Output is provided for one watershed at a time and does not allow one to "quickly" compare results across watersheds or to compare results with water quality output. More guidance is needed for interpreting the significance of elevation frequency distribution diagram with respect to runoff potential.

Although the diagram is interesting, a computation of average watershed slope might be more relevant and useful.

In running the land use, soil, and topographic reports, I sometimes encountered the following message:
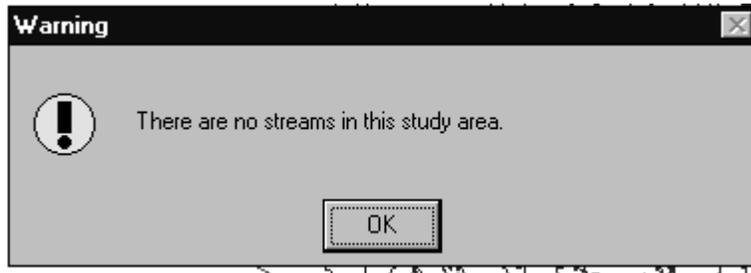
It is not clear how the presence or absence of streams would influence the production of these reports. Specifically, what is meant by "streams"?

BASINS does not provide the ability to generate reports based upon hydrologic soil group (which to some extent integrates over the individual soil properties) or to compute average watershed slopes. These features seem important because they are highly relevant to runoff modeling using widely applied SCS methods.

The reporting functions are potentially useful tools for developing input data sets for nonpoint source assessments. For application of simple export-coefficient models, it would be useful to have a reporting function that produced an ASCII (or DBF) flat file containing the following fields:

1. Subwatershed
2. Land Use Code
3. Area

The total area over records would equal the total area of the watershed. This involves intersecting the land use coverage and watershed delineation.   It seems necessary to have this operate on user-defined subwatersheds, since the USGS HU's are generally too large for typical model applications.

For application of runoff models based upon curve numbers and percent impervious (e.g., SCS methods, P8 (*http://www.shore.net/~wwwalker/ p8/p8.htm*)), it would be useful to have a reporting function that produced an output file with the following fields:

1. Subwatershed

2. Land Use Code
3. Hydrologic Soil Group (A,B,C,D, etc)
4. Area

The total area over all records would equal the total area of the watershed. This involves intersecting the watershed delineation, land use, and soil coverages.

The next level of complexity would consider slope.   In this case, the output file would contain the following fields:

1. Subwatershed
2. Land Use Code
3. Hydrologic Soil Group
4. Slope Class (e.g., 0-1%, 1-2%, etc…)
5. Area

This type of breakdown is used by GISPLM (*http://www.shore.net/~wwwalker/ gisplm/gisplm.htm*).

Each of the above data sets would also contain a record identifier field. A coverage of intersection results would also be produced.   Model output data sets could then be linked to the coverage based upon the record identifier and displayed in ArcView.

It seems that the constraint listed on the bottom of page 7.2-4 (user-defined land uses not available to reporting functions) represents a significant limitation if the reports are to be used to generate input data for models other then NPSM. Even if one accepts the BASINS/ USGS coverage as a base case, there would be no way of evaluating future land use scenarios.

Stream Water Quality Models

Generally, BASINS seems to provide reasonable starting points for applications of QUAL2E, TOXIROUTE, & HSPF.   I had no idea what to expect here, but believe that many users will be expecting BASINS to provide a turnkey operation.   This is not the case and is probably not even a realistic ultimate objective, given the need to shape assessments based upon site-specific conditions (physical conditions, data limitations, parameters of interest, etc.).

It seems, in general, that the requirement to operate on the RF1 reach file is a potentially significant constraint for applying BASINS modelling routines to many watersheds, particularly in regions with higher relief and higher drainage densities.   Extending the applications to the RF3 level and/or to operate on

user-defined reaches would increase applicability. It is not clear in the documentation whether the existing model applications will work with user-defined (imported) reach files. This and other general limitations should be discussed at the beginning of the modeling section (p. 9-1).

It seems appropriate for the documentation to provide a description of exactly what preliminary model formulation steps BASINS can facilitate and what steps are likely to require independent effort on the part of the user. These lists should be provided up-front, rather than sprinkled throughout the documentation on each model. For example, some important independent user efforts in QUAL2E applications are listed in 'Tips' (e.g., 9.1-4 to 9.1-5). These are probably not comprehensive. It seems important to provide a more concise and comprehensive list of such requirements at the beginning of the chapter. In this way, the user will have more realistic expectations.

In testing QUAL2E, I followed the tutorial, more or less.

Page 9.1-2, Item 1, "Reach lengths should not be significantly different". Different from what? Each other?

Page 9.1-4, Item 8, The reference to Screen 9.1.4 does not seem correct. Screen 9.1.4 does not have a 'Run' option to select, as indicated in text.

I found that I was able to get through a QUAL2E session, edit input values, run model, display output, etc. without looking at the QUAL2E documentation, even though I had never seen the QUAL2E windows interface before. The QUAL2E interface and VISUALIZE tool are reasonably intuitive.

The help file for QUAL2E graphics apparently did not come through during the installation.

In applying 'Visualize' to QUAL2E output, different colors are used to represent stream segments with different concentration levels. Are the displayed categories based upon the average or maximum concentrations in each segment? The output theme should be labeled as such.

Nonpoint Source Model

I tried applying NPSM to a single HU in Massachusetts (0170005). The entire watershed was treated as one segment. The watershed/reach map looks like this:

The following warning message was obtained:

The same warning message was encountered when I tried applying NPSM to 01070004. If the streams are not connected, does this mean that there is a problem with the V1 Reach file?   It is not clear what this message means or how it affects results. This message may reflect the presence of a lake segment in the reach network.   The model setup process continued pass this point.

As indicated in the following screen, NPSM apparently lumps all urban land uses into a single category for the purposes of calculating impervious areas:
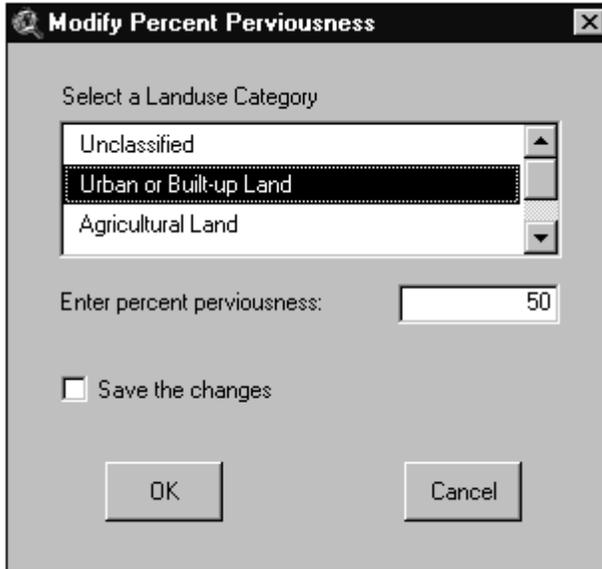
It seems like a great deal of useful information is being ignored by the model at this point, since imperviousness varies significantly with urban land use type (e.g., commercial vs. low density residential). Throwing out this useful information here (which the user can probably get a handle on) while later requiring the user to estimate (or blindly accept) default values for numerous parameters embedded in the simulation seems inappropriate. There does not seem to be much point in running a complex simulation of the watershed if key watershed features are being heavily lumped. The documentation suggests that greater resolution of land use types is possible using an imported land use cover. I was not able to test this, however, because of difficulties with the land use import utility discussed above.

Default values for impervious fraction do not seem at all reasonable (Forest 100%, Agriculture 100%, etc.. Although default values can apparently be changed by the user (10.1-3), better initial values should be provided. What is meant by "impervious". Does this only include urban surfaces, or does it also include rock, extremely poorly drained soils, ice, water, perched wetlands etc.?

The following error message was displayed after the Basins NPSM views were created:

Nonpoint Source Model

ℹ The following reaches from the GIS import file contain incomplete or invalid data:

01080204024

Please complete or edit the missing information using the Reach Editor before running the model.
Most likely the slope, width, depth, or the Manning n are missing.

OK
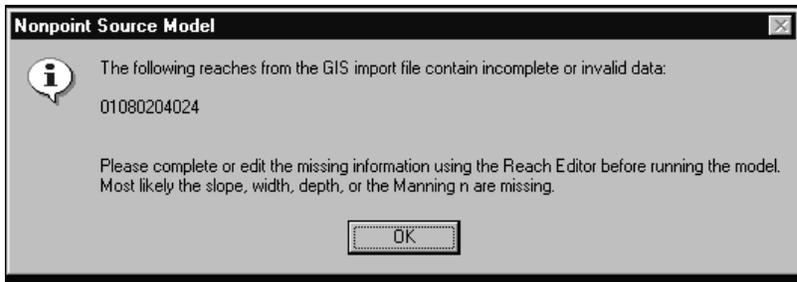
The message should state that the reach editor is accessed from the NPSM menu (as opposed to the BASINS menu).   If the error is in the source Reach file, it is not clear how it can be corrected before creating the NPSM input file.   Is it true that only the output reach network can be edited (not the source ?).   If so, the user seems to be stuck if the source reach network has an error that prevents BASINS from creating a complete NPSM input file for the whole watershed.

The missing data was traced to a lake segment.   It does not seem valid to include a lake segment in NPSM. BASINS documentation provides no guidance on what to do if the basin includes a lake.   Lakes would be frequently encountered when modeling is done on the scale of a cataloguing unit. Should the watershed be subdivided to exclude lakes?

The implied left-to-right editing sequence in the NPSM interface should be given greater emphasis in the documentation (p. 10.3-1). In particular, the requirement to make changes in every button to the right of any edited button should be highlighted (TIP box?).

When the model setup process was complete, the NPSM input file had only 1 reach (representing the lake segment), despite the fact that the source V1 Reach coverage had 28 segments, as shown on the following map:

Apparently, the model setup process fails if the selected watershed contains a lake (?).

I tried running NPSM on a delineated subwatershed that did not contain a lake:
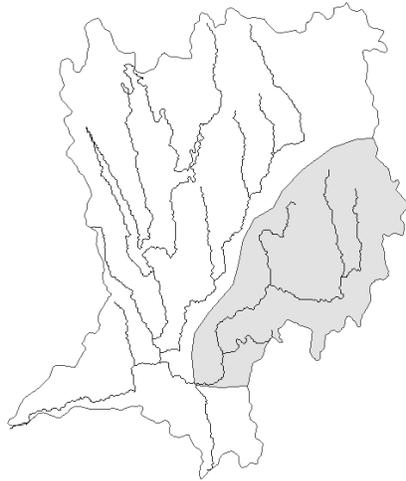
The setup went smoothly, except that the resulting reach network contained only one segment, i.e., it did not represent any of the branches in the Version 1 Reach file:
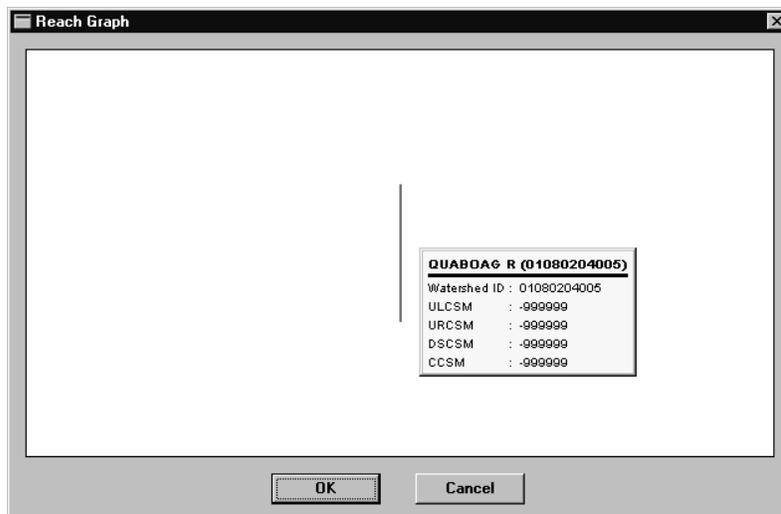
The segment length represented in the NPSM file was 4.9 miles. The total length of the selected subwatershed (measured in ArcView) was ~25 miles.

Apparently, the model only recognized the most downstream segment of the selected reach network, even though the network did not include a lake.

Apparently, BASINS does not automatically identify headwater reaches. This should be more clearly specified (10.4-2)

The Reach Editor would be a lot easier to use if the data entry window could be expanded to show more fields (vs. scrolling left/right to view each input value).

Abbreviations or algebraic terms are frequently used to identify variables for input editing. The definitions of many of these terms (e.g., 'Delta h', 'F-Table') are not intuitive. The screens could be made more explicit. Use English instead of Fortran. A separate table defining each input value could be added to the documentation. This table should indicate which variables are provided by BASINS and which must be independently estimated by the user. Generally, it would be helpful to include a more explicit description of exactly what information is extracted from the GIS layers in the process of creating an NSPM file and what information is not extracted/left to the user.

The TIP on p. 10.4-6 makes reference to "missing elevation and delta h values". Is there a specific code to indicate a missing value? Using 0.0 as a missing value code is not appropriate if 0.0 is a valid entry.

The input interface for the Simulation Time & Met. Data did not seem to accept my choices of weather station. I tried repeatedly selecting 'Worcester Airport':

After closing the window and re-opening it, the selected station always reverted to 'Birch Hill Dam':
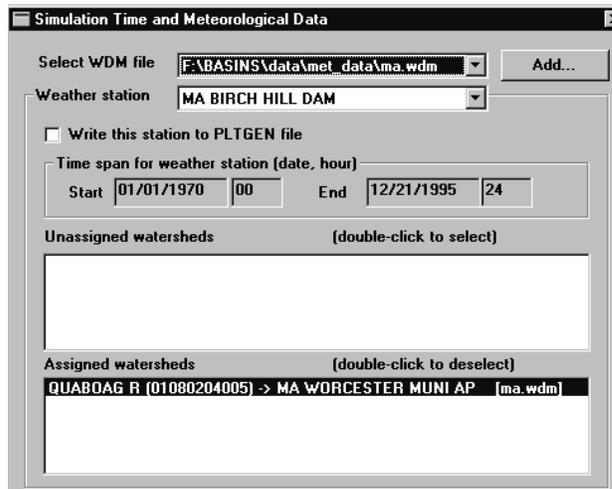
The assignment of the watershed to the Worcester station remained correct, however. Maybe this is working OK (because the assignment is the only factor that matters), but the automatic reversion of the 'Weather Station' input field is confusing.

Is the 'time span for weather station' supposed to change when I select different stations? It never did.

The valid range for the hour field ('HH') on the Simulation Time & Met. Data screen should be specified. A range of 0 to 24 (as indicated in the tutorial box p. 10.5-3) does not seem right, since includes 25 possible values.

More explicit definitions of the fields on the impervious land activity screen should be given:
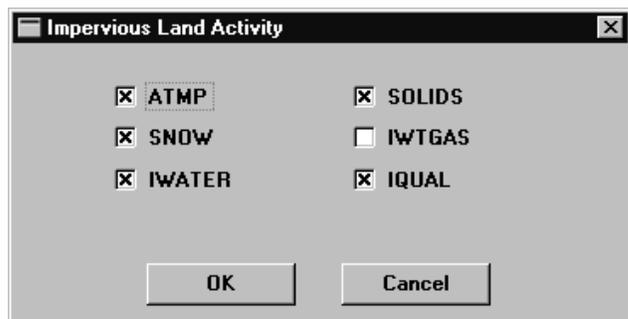
The documentation (p.10.7-1) indicates the need for correspondence between the checked boxes on the pervious and impervious land activity screens.   The correspondence is not obvious in all cases.   For example, what pervious variable corresponds to 'SOLIDS' on the impervious screen? If a 1-1 correspondence is always required, why are there two separate input screens?

Apparantly, ATMP must be selected if SNOW is selected. This should be indicated in the documentation.

Similarly, variables on the reach activity input screen could be identified more clearly (i.e., in English instead of Fortran):
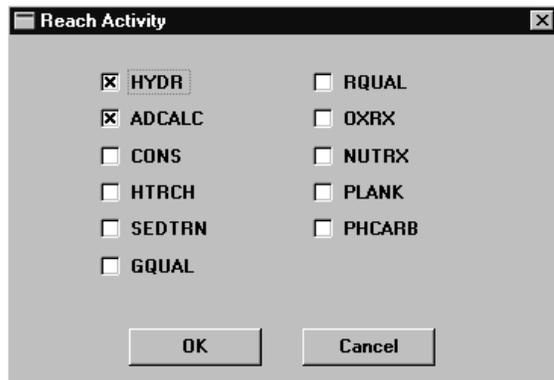
There must be some linkage between the selections on this screen and the selections on the pervious and impervious land activity screens. This linkage is not mentioned in the documentation.

'Gases' is spelled wrong on the 'Pollutant Selection' Screen.

It is not clear why soil properties should be a function of land use.   Soil features and land use features are generally independent.   Future versions of BASINS should consider extracting relevant soil properties from the supplied GIS layer.

Sediment/Solids Distribution

	Land Use Name	Watershed ID	Sand	Silt	Clay
1	Urban or Built-up Land	01080204005	0.00	0.00	0.00
2	Agricultural Land	01080204005	0.00	0.00	0.00
3	Unclassified	01080204005	0.00	0.00	0.00
4	Urban or Built-up Land	01080204005	0.00	0.00	0.00
5	Agricultural Land	01080204005	0.00	0.00	0.00
6	Forest Land	01080204005	0.00	0.00	0.00
7	Barren Land	01080204005	0.00	0.00	0.00

The assignment of default data files for impervious and pervious land units did not appear to work correctly on the test watershed. The watershed included forested areas and an impervious fraction of 0 was specified for forested land use. Yet, all forested land showed up as impervious in the NPSM file.

The NPSM program crashed twice when I attempted to use the Output Manager. The screen enables selection of output from 'all pervious' areas. A corresponding option for all impervious areas seems appropriate. It is not clear from the entry screen or from the documentation how one obtains the combined output from the entire watershed. By trial and error, apparently this is possible when the 'Reach' radio button is clicked.

The variable definitions shown on the bottom of the Output Manager Screen often did not fit into the assigned space (2nd line not legible):

Output Manager

Module: ● Pervious ○ Impervious ○ Reaches

Sum landuse contribution in: 01080204005 (QUABOAG R)

Landuse: Urban or Built-up Land

Output filename: P#_URBAN.005

Print interval: Daily

Output folder: F:\BASINS\modelout\ma

Type: Mean-valued

Select a computed variable to print:

SEDP (2)
SOSEDP
TSP4S (1)
TSP4S (2)
TSP4S (3)
TSP4S (4)
TSP4S (5)
SSP4S (1)
SSP4S (2)
SSP4S (3)
POPHOS (1)
+TRACER

Selected variables:

Fluxes of solution phosphate in the subsoil with deep percolation, from the lower layer to inactive

>>
<<
OK
Cancel

When I ran the model, several error messages flashed by in the DOS window, but I was unable to read them because the window closed automatically. No output files were produced.

When I tried running the model again, it continued running for more than two hours (1 watershed, 1 year simulation) and I had to shut it down manually. I tried deleting all water quality aspects (simulating flow only). Similar results were obtained.

I tried implementing suggestions provided by Tetra Tech, but was unable to get the model running. I was unable to conduct further testing of NPSM because of time limitations.

It is not clear whether there was a problem with the software or erroneous/missing input data. In order to resolve this, it would be helpful to have a complete sample input file supplied with the software. The '/MODELOUT/SAMPLE' directory contains data files, but none are apparently Project files (*.PRJ) for NPSM.

Data Files

Web links to all BASINS data sources should be provided with the documentation and summarized in one place at the BASINS web site.

More details on the landuse coverages supplied with BASINS would be helpful. In particular, what year(s) are they supposed to represent?

The files named "Water Quality Data 79-74" etc. are misnamed. They do not contain data, but data summaries.

Similarly, it seems appropriate to rename the 'Water Quality Station' theme and table to 'Water Quality Summary Station'.

The 'Water Quality Parameter Table" contains water quality criteria (lower, upper values) for some parameters but not others. Most of the criteria are referenced to "Freshwater Aquatic Life" ("Unknown" field in table). The listed references seem dated (most recent 1986). Only one drinking water criterion is supplied (Total Nonfilterable Solids), despite the fact that criteria/standards exist for many of the other listed parameters. This does not seem to be consistent. The table should either be consistent/thorough or nonexistent.

Criteria are also listed in the "Water Quality Criteria Table". The reason for criteria appearing in two separate tables is unclear.   Are the values consistent? Is the table intended to refect "official" criteria that must be used in the TMDL process?   The listed Drinking water MCL for nitrate+nitrite N is 10000 mg/l; this should be 10 mg/l, unless the MCL has changed recently.   The listed acute toxicity criterion for unionized ammonia is 93 mg/liter as N!   The listed acute freshwater criterion for total phosphorus is 1 mg/liter. There may be other errors in the criterion table (I did not check them all).   What is the meaning of a criterion value of 0 (vs. 'unknown') in this table?

The general structure of the water quality summary files limits the scope and usefulness of BASINS assessment tools, graphic output, and tabular output. The reasons for adopting this particular structure probably reflect tradeoffs between the desire to provide useful information and the need to limit data file size etc.   Assessments are limited to the particular 5-year intervals chosen for the data summaries.   A more flexible approach would be to supply relevant data in observation form (vs. summary) and allow the user more control over the data selected for analysis. Summaries would be generated on the user's computer using data selected by the user based upon date, season, depth, station type, and/or other relevant criteria.   This would involve a larger up-front investment in downloading raw data.

If the assessment tools worked from raw data (vs. summaries), they could also be used to analyze user-supplied data.   Although BASINS provides a utility for augmenting the water quality observation tables, these tables are not accessed by the assessment tools (only NPSM ?).

In the test watersheds which I downloaded, only a very small fraction of the total available water quality data were reflected in the data summary files.   In reality, much of the information being used in load allocations is not getting through to STORET (at least in a timely fashion).   Many agencies (especially, water utilities) do not participate. The capability for analysis of locally-supplied data seems important.

It is not clear how the 5-year data summaries were generated at stations that were sampled at more than one depth. Were all depths used? Were the values averaged by date first and then summarized across dates?

It is not clear what criteria are applied to select data for inclusion in the water quality observation tables provided with BASINS. Very few observations came across in the test watersheds.   These files use a 6-character date field. What about Y2000 compatibility?

A greater emphasis on supplying the data in raw form (vs. summary) would also allow possibilities for other types of applications, including:

- trend analysis (e.g. Seasonal Kendall Tests)

- load calculations (integrating measured flow & concentration data)

- more meaningful data displays (box plots, etc.).

These external applications could presumably be developed to be accessed from within the BASINS shell. The raw data would also be more applicable in analyses and modeling conducted outside of BASINS (which would probably more often than not represent cornerstones of final TMDL efforts, given frequent needs to tailor the analyses to site-specific conditions).

One potential application would be to use the BASINS interface as a framework for structuring and submitting STORET retrievals. The initial emphasis would be on downloading station indexes and observation counts, date ranges, etc., rather than data values. The software and interface could then be used to facilitate data selection, create STORET data retrieval instructions, submit data request, and download results. Output data files obtained from STORET would preferably in a format that is useful for further processing, such as DBF or Excel files. I have found the output format options currently available from STORET (test, comma delimited) to be difficult to work with and unreliable. For example, importing a comma delimited file into an Excel workbook does not provide proper alignment of columns if the station code happens to contain a comma or certain other characters.

Similarly, the BASINS framework also seems adaptable and potentially useful as a means for identifying and retrieving USGS daily flow data. It would not be a giant step to include daily flow data from selected stations along with other files supplied by BASINS.

Toxic Release Inventory data files are provided. Site names and locations are given, but where is the information describing the substances involved? Is this available electronically?

General

BASINS represents a significant product which should be useful in TMDL and other applications. In its current form, the package is most useful for quickly acquiring and reviewing basic data inventories and coverages that might be useful in the TMDL process. There still exists a significant gap, however,

Supplement to "Review of Basins 2.0"

Questions & Answers

W. Walker for Cadmus Group, Inc. & USEPA

September 24, 1998

This document has been prepared as a supplement to my September 14 review to assist Cadmus & EPA in extracting answers to specific questions in the charge to reviewers. The review was structured around BASINS components, rather the charge outline.

Major sections of the charge and review generally correspond as follows:

Charge	Review	Pages
A. Overall Concept…	General	30-32
B. BASINS Data Layers	Data Files	28-30
C. Assessment Tools	Assessment Tools	7-9
D. Utilities	Utilities	9-14
E. Watershed Char. Reports	Reports	14-18
F. NPSM	Nonpoint Source Model	19-28
G. QUAL2E	Stream Water Quality Models	18-19
H. TOXIROUTE	(not reviewed)	
I. User's Manual	no specific section, comments in other sections	
J. Planned Enhancements	General	31-32
-----	Installation	3-7

Questions are partially repeated in italics.

Answers refer to the attached annotated copy of the review.

A-30 = bracketed section "A" starting on page 30

Additional elaboration is provided, were appropriate.

A. Overall concept and design

see "General" section, pp. 30-32

Is BASINS, taken as a whole, scientifically sound to make credible TMDL decisions in the nation's urban, rural, and mixed land use situations?

BASINS provides a sound framework for acquiring and analyzing basic data and is useful in the TMDL process. BASINS will not provide answers or make decisions, however.

see A-30,   B-32

A turnkey solution would provide all the answers right out of the box without user intervention. That is, the user would define the region/watershed/reach of interest and the software would retrieve all of the relevant data and produce a valid, calibrated model that could be used in TMDL's. BASINS is not there yet (nor is any other software that I know of). It does a lot of the initial ground work and saves time ("jump start"), but it is likely that the user will have to locate and integrate substantial additional local data and exercise considerable judgment in order to develop a valid model. The process is still largely subjective. I am not criticizing BASINS for not being a turnkey solution. I am suggesting, however, that these limitations be more clearly stated in the documentation and "advertising" (web site, etc.), so that users (or potential users) have more realistic expectations. Otherwise, poor management decisions may be made because some users believe BASINS answers (because it is EPA's "model") without exercising any of their own judgment, introducing site-specific data, or using models that may be more realistic and appropriate than those provided with BASINS.

Can BASINS be used to perform TMDLs under a variety of sources and pollutant discharge situations to address a significant number of the' thousands of TMDLs expected in the next 10 years? The panel should keep in mind that the constituent models do not yet include a lake or estuary model.

BASINS is useful in the TMDL development process, subject to limitations discussed above. In its current form, the software is most useful for data compilation and preliminary analysis.

Lack of a lake/reservoir model is a significant gap. (C-32)

Are the individual components of BASINS adequately comprehensive to meet EPA's management objectives for controlling point and nonpoint source pollution…? Are the models chosen adequate to perform TMDL's for these pollutants in non-tidally influenced rivers and streams?

QUAL2E is comprehensive for addressing low-flow problems in streams and shallow impoundments, most often related to point sources.

NPSM is too general, too complex, and not within the grasp of most potential users. Simpler tools should be provided to address common problems most often related to nonpoint sources, such as sedimentation and eutrophication. (D-31).

Additional reporting functions and/or direct model integrations are suggested to facilitate applications of other nonpoint source models (BB-17, D-31). This will make BASINS

palatable to a wider user community and provide greater flexibility for application of alternative models.

The basic "building blocks" of BASINS are existing GIS software (ArcView 3.0a), nationally available environmental databases maintained by EPA and other federal agencies, and established water quality and watershed models. The BASINS development process involved the creation of interfaces to provide a seamless framework and automated extraction functions between the databases and models, in order to facilitate preparation of model input files and output presentation. Is it scientifically valid to integrate data, GIS, and models in this fashion? Does the overall design of the linkages between the GIS, data layers, reports, models, and presentation of results follow a logical pattern?

The overall integration concept is good, but BASINS would be more effective if simpler models were also integrated. (D-31).

Compatibility with ArcView 3.1 seems important (U-3).

BASINS comes bundled with national databases and default data sets to act as a starting point for the users analysis, however the user is given the option of overriding these dam Does the system strike an adequate balance between the need to provide the user with data and the recognition of the site specificity often required?

The data bases are valid starting points. Most applications will require substantial input of site-specific data. There is room for improvement in the capability for entering and utilizing new data:
I had difficulties entering a new land use coverage (F-12) and appending new water quality data. (G-13).

Potentially useful reporting functions do not work with imported coverages (H-14, J-18).

Apparently, even if the user is successful in augmenting the water quality observation table (I was not, G-13), the data cannot be accessed by BASINS assessment tools (O-29).

Based upon experience with test cases, an ability to edit reach files or to enter new ones also seems important (I-18, K-22). The distributed reach files may be in error or not contain sufficient detail. There does not seem to be any remedy. The extraction process (automated model setup) reaches a dead end (requiring manual input of data directly to model(s)) if the distributed reach files are not adequate for a particular watershed.

B. BASINS data layers

see "Data Files" section, pp. 28-30

Are the data layers provided in BASINS appropriate and adequate for the environmental analysis requirements for TMDLs, i.e. large and small scale characterization of the watershed setting and its primary pollution sources, and nonpoint and point source modeling?

They are appropriate, but not necessarily adequate because of the probable need to supplement them with site-specific data derived from other sources, as discussed above.

Data layers for hydrologic soil groups and slope would be critical to hydrologic and nonpoint source modeling (L-17).

Structure of water quality files limits potential uses (O-29).

Test watersheds contained only a small fraction of the available water quality data (P-29). Some of this reflected the fact that the collecting agencies did not work with STORET. The Michigan coverage I downloaded, however, did not reflect massive amounts of data that I know are in STORET. The criteria for including or excluding specific water quality data from STORET are not clear (Q-29).

Are the data dictionary and meta-data adequate?

Suggest providing links to data sources and more detail on supplied land use coverages (M-28).

See comments on Water Quality Parameter Table (N-28).

An online version of the data dictionary would be useful (help file, web page…).

C. Assessment Tools

BASINS provides several tools (Target, Asses, and Data Mining) designed to assess instream water quality conditions and point source discharges at a regional Watershed and stream segment level. For each of these tools the reviewers should comment on its user friendliness, and its usefulness in the analysis of watersheds and preparation of reports and records for TMDLs.

see "Assessment Tools" section, pp. 7-9.

D. Utilities

see "Utilities" pp. 9-14.

Data management tools...

a. *Watershed delineation*   (R-9)
b. *Import*   (F-12)

c. *Land Use Reclassification*   (F-12)
d. *Digital Elevation Model Reclassification*   (S-13)
e. *Look-up Tables*   (N-28)
f. *Water Quality Observation Management*   (T-13)

In particular the panel should comment on whether the utilities are helpful in allowing the user to customize data layers to address site specific conditions. The panel may also wish to suggest additional utilities which could address commonly encountered problems.

See above comments in Section A, Data.

A utility to edit & import reach files seems appropriate (I-18, K-22).

More guidance and/or utilities are needed on the "projection" concept in building projects (V-5, U-6).

Suggest potential utilities/reports that use BASINS as a framework for computing loads, performing trend analyses, submitting STORET data requests, and retrieving USGS streamflow data (EE-30).

E. Watershed Characterization Reports

see "Reports", pp. 14-18.

a. *Point Source Inventory Report*   (W-14)
b. *Water Quality Summary Report*   (W-14)
c. *Landuse Distribution Report*   (H-14, X-14)
d. *Toxic Air Emission Report*   not tested
e. *State Soil Characteristic Report* (Y-15)
f. *Watershed Topographic Report*   (Z-16)

For each report, the panel should evaluate its design and usefulness to the user in developing watershed characterization and analyses and an administrative record.

Frequently encountered warning message ("no streams…") in generating reports (AA-17).

Several suggestions for additional reports that would enable Basins use with other nonpoint source models (BB-17).

F. NPSM

see "NonPoint Source Model", pp 19-28.

As discussed in my review (p 28), I was unable to produce a successful run with NPSM.

The problems may partially reflect my lack of previous experience with the model. The model has too steep a learning curve to allow me to become familiar enough with it to provide answers to most of the specified questions within the time frame allotted for this review.

Does NPSM correctly extract data from the GIS data layers?

There is an apparent problem when the watershed contains a lake (CC-20).

The software did not seem to correctly extract reach configuration (K-22, DD-22).

The assignment of default data files for impervious and pervious land units did not appear to work correctly (GG-27).

Lumping of all urban land uses into one category for computing impervious area tosses out significant information (FF-20). Impervious area is the most important factor for predicting urban nonpoint loads.

Are the data layers chosen as model input appropriate?

The model seems to ignore important features (hydrologic soil group, slope, erodibility) that control runoff generation and erosion potential.

BASINS provides a default data set which allows NPSM to run; the user can then use these data as a starting point for calibration of the model for the watershed being studied. The HSPF Data Dictionary in the user's guide provides definition of each input parameter, maximum and minimum values, and the default Value. The panel should evaluate, whether this a reasonable approach, and whether it strikes an adequate balance between the need for user support and the recognition of the site specificity and user knowledge required.

See general comments on NPSM (D-31).

G. QUAL2E

see "Stream Water Quality Models" pp. 18-19.

Is the Windows interface an accurate reflection of the input data requirements of QUAL2E?

While I did not spend much time with it, the interface seems to provide access the basic input data groups.

Are the data layers chosen as model input appropriate?

Capability to operate on RF3 (or user-defined) reach file would be useful.

Does QUAL2A correctly extract data from the GIS data layers?

In some of my tests, the reach network was not extracted.

H. TOXIROUTE

Not reviewed because of time limitations.

I. User's Manual

Several comments on the documentation are distributed throughout the review.

To enable users to develop realistic expectations, the documentation should contain a clear discussion of limitations (A-30), including explicit descriptions of which model inputs are provided by basins and which must be independently supplied/estimated by the user (FF-19, GG-24). The documentation should not "over-sell" the software, but convey sufficient information allow the user (or potential user) to develop realistic expectations. This will reduce levels of frustration and misuse.

More guidance on the "projection" concept is needed to facilitate creation of projects and importing of coverages (U-6).

J. Planned Enhancements

Although secondary to the panel's charge of review of the current capabilities of BASINS version 2.0, the panel may wish to provide suggestions on future directions for BASINS.

Reach Input/Editor (I-18, K-22)
Reports on Watershed Features Controlling Runoff Potential (L-17)
Reports to Generate Input for other Nonpoint models (BB-17)
Better Access to Raw Water Quality Data (vs. Summaries) (O-29)
Trend Analyses, Load Calculations (EE-30)
Framework for STORET/USGS data retrievals (EE-30)
Simpler Nonpoint Models (D-31, H-14)
Lake Models (C-32)
ArcView 3.1 compatibility (U-3)
More Capability for Selecting & Screening WQ Data to be Analyzed (GG-8)

between the data compilations and inventories that are provided by BASINS and the development of a valid modeling framework for TMDL's. This gap partially reflects:

- limitations in the coverages & tables extracted from nationwide data bases;

- difficulties involved in specifying a "general" procedure that is applicable to most watersheds without considerable site-specific adjustments;

- limited selection of models; in particular, NPSM/HSPF is too complex, dependent on numerous buried assumptions, and has too steep a learning curve to be practical as the only BASINS option for modeling nonpoint loads.

The last problem is consistent with EPA's historical tendency to promote development of models that are unwieldy and overly complex, given limitations in data, time, and expertise that are typically encountered by water-quality management agencies charged with developing TMDL's. In developing TMDL's, I believe that most analysts would rather rely on gross assumptions that can be directly supported (e.g., export coefficients developed from regional measurements) than on numerous assumptions (e.g., kinetic, stoichiometric coefficients) that are buried in a complex model, unfamiliar, highly uncertain, and not directly measureable. There are probably only a handful of people who claim to be fluent with HSPF and a smaller number who can use the model to produce valid results. Unless the firm is planning to expand considerably, hiring Tetra Tech to develop TMDL's is an option only for a small fraction of TMDL efforts.

For the above reasons, it seems important to provide alternative, screening-level and planning-level models for use in non-point source assessments. Some options are described in "Compendium of Watershed-Scale Models for TMDL Development" (EPA841R94002). One level of analysis would be based upon export coefficients (unit area loads) and/or runoff coefficients related to land use. Coefficient values could be user-defined and/or extracted from regional data summaries (e.g., nationwide maps of non-point-source nutrient export coefficients and stream concentrations developed by Omernik of EPA Corvallis based upon data from the USEPA National Eutrophication Survey; these maps might be supplied as coverages). Equations of the type shown in slide 101-18 of the course handout notes are often practical and appropriate, but are apparently not implemented in the current version of BASINS. Schuler's 'Simple Method' is a valid and practical tool for estimating nonpoint loads from urban and mixed watersheds.

Based upon the list of "Planned Enhancements" (Charge to Reviewers, Page 7), future directions appear to be towards models that are more complex and less practical. These enhancements may widen, rather than bridge, the existing gaps between the data accessible within the BASINS framework and the modeling efforts required to develop TMDL's.

The absence of a framework for evaluating lakes (either in BASINS 2.0 or in the Planned Enhancements listed in the Charge to Reviewers) is unfortunate, considering the widespread impacts of point and nonpoint sources on lakes, apparent needs for lake TMDL's, and the fact that simple, practical models are available for these types of assessments. Impacts of land use on phosphorus export and lake eutrophication are fairly widespread, well-defined, and quantifiable using methods that are simpler and probably more reliable than HSPF.

In its current form, BASINS provides a jump-start in TMDL development by supplying and manipulating relevant data and coverages that may be useful for implementing models supplied with BASINS, as well as others. For those models that have been implemented, more explicit descriptions are needed of exactly which input variables are extracted by BASINS and which variables must be estimated by the user. This will enable users (and potential users) to develop more realistic expectations; many may have the mistaken impression that BASINS provides a turnkey modeling solution.