



# **Water Security Initiative: Event Detection, Deployment, Integration and Evaluation System (EDDIES), Version 4.4 User Reference Guide**

## Disclaimer

The Water Security Division of the Office of Ground Water and Drinking Water has reviewed and approved this document for publication. This document does not impose legally binding requirements on any party. The information in this document is intended solely to recommend or suggest and does not imply any requirements. Neither the U.S. Government nor any of its employees, contractors or their employees make any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use of any information, product or process discussed in this document, or represents that its use by such party would not infringe on privately owned rights. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Questions concerning this document should be addressed to:

Katie Umberg

EPA Water Security Division 26 West Martin Luther King Drive Mail Code 140 Cincinnati, OH 45268  
(513) 569-7925

[Umberg.katie@epa.gov](mailto:Umberg.katie@epa.gov)

or

Steve Allgeier

EPA Water Security Division 26 West Martin Luther King Drive Mail Code 140 Cincinnati, OH 45268  
(513) 569-7131

[Allgeier.Steve@epa.gov](mailto:Allgeier.Steve@epa.gov)

## Acknowledgements

The Water Security Division of the Office of Ground Water and Drinking Water would like to recognize the following individuals and organizations for their assistance, contributions and review during the development of this document.

- Jim Alair, New York City Department of Environmental Protection
- David Hart, Sandia National Laboratories
- Yeongho Lee, Greater Cincinnati Water Works

# Table of Contents

<b>SECTION 1.0: INTRODUCTION</b> .....	<b>1</b>
1.1 ONLINE WATER QUALITY MONITORING AND EVENT DETECTION OVERVIEW .....	1
1.2 EDS OUTPUTS .....	2
<b>SECTION 2.0: EDDIES 4.4 OVERVIEW</b> .....	<b>4</b>
2.1 EDDIES CAPABILITIES .....	4
2.2 THE EDDIES INTERFACE .....	5
<b>SECTION 3.0: EDDIES INSTALLATION AND CONFIGURATION</b> .....	<b>7</b>
3.1 ORACLE DATABASE INSTALLATION AND SETUP .....	7
3.2 EDDIES APPLICATION INSTALLATION AND CONFIGURATION .....	7
3.2.1 <i>EDDIES Installation</i> .....	7
3.2.2 <i>Oracle Database Setup</i> .....	8
3.2.3 <i>Launching EDDIES and Connecting to the Oracle Database</i> .....	9
<b>SECTION 4.0: MONITORING SYSTEM AND UTILITY DATA MANAGEMENT</b> .....	<b>11</b>
4.1 MONITORING SYSTEM DATA .....	11
4.1.1 <i>Parameter Types</i> .....	12
4.1.2 <i>Parameters</i> .....	16
4.1.3 <i>Locations</i> .....	19
4.1.4 <i>Parameter Locations</i> .....	21
4.2 IMPORTED DATA .....	24
4.2.1 <i>Viewing Imported Data</i> .....	24
4.2.2 <i>Adding Utility Data</i> .....	24
4.2.3 <i>Editing Imported Data</i> .....	25
<b>SECTION 5.0: EDS SETUP</b> .....	<b>26</b>
5.1 DATA ANALYSIS OPTIONS .....	26
5.1.1 <i>EDDIES-compatible EDS</i> .....	26
5.1.2 <i>External EDS</i> .....	26
5.1.3 <i>Setpoint EDS</i> .....	27
5.2 EDS INSTALLATION AND REGISTRATION .....	27
5.2.1 <i>EDS Software Installation and Configuration</i> .....	27
5.2.2 <i>EDS User Account</i> .....	27
5.2.3 <i>EDS Registration</i> .....	28
5.2.4 <i>EDS Variables</i> .....	30
5.2.5 <i>EDS Configuration</i> .....	33
<b>SECTION 6.0: CREATION AND EXECUTION OF BATCHES</b> .....	<b>37</b>
6.1 CREATE A BATCH OF RUNS .....	37
6.1.1 <i>Batch Properties</i> .....	37
6.1.2 <i>Methods for Defining a Batch</i> .....	38
6.2 EXECUTE A BATCH .....	40
6.2.1 <i>Batch Variables</i> .....	41
6.2.2 <i>Executing a Batch</i> .....	43
<b>SECTION 7.0: EXPORT AND ANALYSIS OF RESULTS</b> .....	<b>47</b>
7.1 SELECTING RUNS TO EXPORT .....	47
7.1.1 <i>Filtering</i> .....	48
7.2 EXPORT TYPES .....	49
7.2.1 <i>Alerts Export</i> .....	49
7.2.2 <i>Results and Data Export</i> .....	50
7.2.3 <i>Analysis Data Export</i> .....	52
7.2.4 <i>Setpoint Sensitivity Export</i> .....	52

<b>GLOSSARY .....</b>	<b>53</b>
<b>APPENDIX A: IMPORT FILE FORMATS.....</b>	<b>55</b>
A.1    PARAMETER TYPE IMPORT.....	55
A.2    PARAMETER IMPORT.....	56
A.3    LOCATION IMPORT.....	56
A.4    LOCATION PARAMETER IMPORT .....	57
A.5    UTILITY DATA IMPORT .....	57
A.5.1 <i>Time-series Format</i> .....	57
A.5.2 <i>Parameter-based Format</i> .....	58
A.6    EDS REGISTRATION IMPORT.....	59
A.6.1 <i>EDS Import</i> .....	59
A.6.2 <i>EDS Property Import</i> .....	59
A.7    EDS VARIABLE IMPORT.....	60
A.8    EDS CONFIGURATION IMPORT.....	60
A.9    CONTAMINATION EVENT LIBRARIES IMPORT.....	61
A.9.1 <i>Contaminant Import</i> .....	61
A.9.2 <i>Event Profile Import</i> .....	62
A.10   BATCH AND RUN IMPORT .....	62
A.11   EDS RESULTS IMPORT.....	63
<b>APPENDIX B: DATASET GENERATION.....</b>	<b>64</b>
B.1    APPLICATION OF LOCATION, TIME PERIOD AND POLLING INTERVAL IN DATASET CREATION .....	64
B.1.1 <i>Batch Location</i> .....	64
B.1.2 <i>Batch Time Period</i> .....	64
B.1.3 <i>Batch Polling Interval</i> .....	65
B.2    SIMULATION OF CONTAMINATION EVENTS.....	66
B.2.1 <i>Contaminant</i> .....	66
B.2.2 <i>Peak Contaminant Concentration</i> .....	68
B.2.3 <i>Event Profile</i> .....	69
B.2.4 <i>Event Start Time</i> .....	71
B.2.5 <i>Simulated Contamination Event Example Calculation</i> .....	72
B.3    CONTAMINANT AND EVENT PROFILE LIBRARIES .....	74
B.3.1 <i>Contaminant Library</i> .....	75
B.3.2 <i>Event Profile Library</i> .....	77
<b>APPENDIX C: EXPORT FILE FORMATS.....</b>	<b>80</b>
C.1    ‘TEST DATASET’ EXPORT .....	80
C.1.1 <i>Time-series Format</i> .....	80
C.1.2 <i>Parameter-based Format</i> .....	80
C.2    ‘LOG RESULTS AND DATA’ EXPORT .....	81
C.3    ‘ALERTS’ EXPORT.....	81
C.4    EXPORT RESULTS AND DATA.....	82
C.5    ANALYSIS DATA EXPORT .....	83
C.6    SETPOINT SENSITIVITY ANALYSIS EXPORT.....	87
<b>APPENDIX D: KEY TERMS AND ANALYSIS METHODOLOGY .....</b>	<b>88</b>
D.1    EDS OUTPUT .....	88
D.2    KEY TERMS AND METRICS.....	89
D.2.1 <i>Key Terms</i> .....	89
D.2.2 <i>Data Analysis Metrics</i> .....	91
D.2.3 <i>Analysis Settings</i> .....	97
D.3    EXAMPLE CALCULATIONS .....	97
D.3.1 <i>Example Scenario</i> .....	98
D.3.2 <i>Calculation of Metrics Independent of Discrimination Threshold</i> .....	99
D.3.3 <i>Alert Metrics</i> .....	100

D.3.4	<i>Detection Metrics</i> .....	107
<b>APPENDIX E: SETPOINT ANALYSIS .....</b>		<b>111</b>
E.1	SETPOINT ALGORITHM.....	111
E.1.1	<i>Creating a Batch Using the Setpoint Algorithm</i> .....	112
E.1.2	<i>Analyzing Results from the Setpoint Algorithm</i> .....	113
E.2	SETPOINT SENSITIVITY ANALYSIS .....	114
E.2.1	<i>Using the Setpoint Sensitivity Analysis Tool</i> .....	114
E.2.2	<i>Analyzing Results from the Setpoint Sensitivity Analysis Tool</i> .....	115
<b>APPENDIX F: DATABASE SCRIPTS .....</b>		<b>119</b>
F.1	DEPLOY DATABASE SCRIPT .....	119
F.2	DEPLOY EDS DATABASE SCRIPT .....	119
F.3	EXPORT DATABASE SCRIPT .....	119
F.4	IMPORT DATABASE SCRIPT .....	120
<b>APPENDIX G: EDDIES KEYBOARD SHORTCUTS.....</b>		<b>121</b>
G.1	MAIN MENU BAR.....	121
G.2	NAVIGATION WITHIN AN EDDIES TAB SCREEN .....	121
<b>APPENDIX H: COMMON ISSUES .....</b>		<b>123</b>

## List of Tables

Table 4-1. Parameter Types Fields .....	12
Table 4-2. Parameter Fields .....	16
Table 4-3. Location Fields .....	19
Table 5-1. EDS Registration Fields .....	28
Table 5-2. EDS Variable Fields .....	31
Table 7-1. Sample ‘Alerts’ Export.....	49
Table 7-2. Sample ‘Results and Data’ Export .....	51
Table A-1. Sample Parameter Type Import CSV File .....	55
Table A-2. Sample Parameter Import CSV File .....	56
Table A-3. Sample Location Import CSV File .....	56
Table A-4. Sample Location Parameter Import CSV File .....	57
Table A-5. Sample Baseline Data Import CSV File in Time-series Format.....	58
Table A-6. Sample Baseline Data Import CSV File in Parameter-based Format .....	58
Table A-7. Sample EDS Import CSV File.....	59
Table A-8. Sample EDS Property Import CSV File .....	60
Table A-9. Sample EDS Variable Import CSV File .....	60
Table A-10. Sample EDS Configuration Import CSV File .....	61
Table A-11. Sample Contaminant Import CSV File.....	61
Table A-12. Sample Event Profile Import CSV File .....	62
Table A-13. Sample Batch and Run Import CSV File.....	63
Table A-14. Sample EDS Results Import CSV File .....	63
Table B-1. Example A_PH Baseline Data .....	65
Table B-2. Example A_PH Baseline Data Transformed to a 2-Minute Polling Interval .....	65
Table B-3. Example A_PH Baseline Data Transformed to a 10-Minute Polling Interval .....	65
Table B-4. Example A_CL2 Baseline Data .....	66
Table B-5. Resulting Data in the Baseline Dataset Using a 10-Minute Polling Interval .....	66
Table B-6. Default Contaminants in Terms of Contaminant Concentration.....	68
Table B-7. Example Baseline Dataset .....	73
Table B-8. Sample Event Profile .....	73
Table B-9. Chlorine Changes Resulting from Simulated Event .....	74
Table B-10. Simulated Event Dataset.....	74
Table C-1. Sample Imported Data Export File in Time-series Format.....	80
Table C-2. Sample Imported Data Export File in Parameter-based Format .....	81
Table C-3. Sample Log Results and Sample Data File .....	81
Table C-4. Sample ‘Alerts’ Export.....	82
Table C-5. Sample Results with Sample Data Export File .....	83
Table C-6. Sample Basic Analysis Data Export File .....	84
Table C-7. Sample Basic Analysis Data Export File .....	85
Table C-8. Sample Setpoint Low Export File.....	87
Table D-1. “Short hand” used in Section D.3 .....	98
Table D-2. Example Baseline Run: “Eg_Base”.....	98
Table D-3. Example Simulated Event 1: “Eg_Ev1” .....	99
Table D-4. Example Simulated Event 2: “Eg_Ev2” .....	99
Table D-5. Eg_Ev1 Net Response Calculations .....	100
Table D-6. Eg_Ev2 Net Response Calculations .....	100
Table D-7. Alerting Timesteps for Eg_Base.....	102
Table D-8. Alerting Timesteps for Eg_Ev1 .....	102
Table D-9. Alerting Timesteps for Eg_Ev2.....	103
Table D-10. Eg_Ev1 Alerting Timestep Calculations based on Net Response .....	103
Table D-11. Eg_Ev2 Alerting Timestep Calculations based on Net Response .....	103
Table D-12. Normal Timestep Classification for Eg_Base .....	105
Table D-13. Eg_Base Alert Determination for a Discrimination Threshold of 0.75 .....	106
Table D-14. Eg_Ev1 Alert Determination for a Discrimination Threshold of 0.75 .....	106
Table D-15. Determination of Events Detected for three Required Ratios .....	107

Table D-16. Determination of Time to Detect when Required Ratio = 0.75 .....	109
Table D-17. Trigger Parameters for Eg_Ev1 .....	110
Table E-1. Example Data and Setpoint Algorithm Output .....	112
Table E-2. Sample High 'Setpoint Sensitivity Analysis' Export File.....	116
Table E-3. Sample Low 'Setpoint Sensitivity Analysis' Export File .....	117

## List of Figures

Figure 1-1. Example EDS Output during a Simulated Event .....	3
Figure 2-1. EDDIES Tabs.....	5
Figure 3-1. DOS Window Opened By Deploy Script.....	8
Figure 3-2. 'Database Selection' Window.....	10
Figure 4-1. 'View' Menu.....	13
Figure 4-2. 'Parameter Types' Viewing Window.....	13
Figure 4-3. 'Add' Menu.....	14
Figure 4-4. Parameter Type Editor.....	14
Figure 4-5. 'Import Manager' Tab.....	15
Figure 4-6. 'Import Monitoring System Data' Menu .....	15
Figure 4-7. Parameter Types CSV Import Confirmation.....	16
Figure 4-8. 'Parameters' Viewing Window.....	17
Figure 4-9. 'Location Manager' Tab .....	17
Figure 4-10. Add a Parameter.....	18
Figure 4-11. Edit a Parameter .....	19
Figure 4-12. View Locations .....	20
Figure 4-13. 'Add a Location' Window .....	20
Figure 4-14. Filter Criteria.....	22
Figure 4-15. Figure 4-9 with TOC Filter Applied .....	23
Figure 4-16. View Available Imported Data .....	24
Figure 4-17. Import Utility Data (choose format) .....	25
Figure 4-18. Time-Series Based Import Format Verification .....	25
Figure 4-19. Parameter-based Import Format Verification .....	25
Figure 5-1. Select an EDS for Viewing .....	29
Figure 5-2. Viewing EDS Registration Information.....	29
Figure 5-3. Select EDS for Configuration .....	32
Figure 5-4. View EDS Variables .....	32
Figure 5-5. 'EDS Variable Editor' Window .....	33
Figure 5-6. Viewing Configuration Variables for a Given EDS Configuration .....	34
Figure 6-1. Batch Manager.....	39
Figure 6-2. Analysis Data Import Menu .....	40
Figure 6-3. Enter Batch Information.....	40
Figure 6-4. Launch Manager .....	41
Figure 6-5. 'Edit' Menu.....	42
Figure 6-6. Batch Processing Progress .....	44
Figure 6-7. Batch Info .....	46
Figure 7-1. 'Export Manager' Tab.....	47
Figure 7-2. 'Export Manager' Tab Filter .....	48
Figure 7-3. Analysis Settings.....	50
Figure 7-4. Alert Identification Method .....	50
Figure B-1. Example Simulated Event Varying by Contaminant.....	67
Figure B-2. Example Simulated Event with Low and High Peak Contaminant Concentrations .....	69
Figure B-3. Example Contamination Event with Flat and Steep Event Profiles .....	70
Figure B-4. Default Event Profiles A, B, C, D, E and F.....	71
Figure B-5. Example Simulated Contamination Event with Varying Start Times .....	72
Figure B-6. Contaminant Library Window.....	75
Figure B-7. Simulated Event Using Constant Terms in Reaction Factor .....	76
Figure B-8. 'Add Parameters and Reaction Factors' Window .....	77
Figure B-9. View Event Profile Library .....	78
Figure B-10. 'Add an Event Profile Time Step' Window .....	79
Figure D-1. Example EDS Output.....	89
Figure D-2. Analysis Settings.....	97
Figure F-1. Export Progress.....	120
Figure G-1. Main Menu Bar with Shortcuts Active .....	121

## Section 1.0: Introduction

This document is a user reference guide for the Event Detection, Deployment, Integration and Evaluation System (EDDIES), version 4.4. EDDIES 4.4 is an off-line tool developed by the EPA to facilitate implementation of online water quality monitoring. It was initiated for, and has been enhanced based on, the needs of the EPA's Water Security Initiative pilot utilities as they select and implement event detection. Anticipated users of EDDIES in the future include water utilities, contractors, EDS developers and researchers. It generates datasets for EDS analysis, using utility data and simulating contamination events. It then analyzes the output produced by the EDS to evaluate EDS performance.

The document is organized as follows:

- Section 1: Introduction and overview of EDSs.
- Section 2: Overview of the functionality and use of the EDDIES tool.
- Sections 3 – 5: Instructions for initial setup of EDDIES. This includes installing and configuring EDDIES, as well as entering system information and uploading data to be used for evaluation. These sections are organized in the order that the steps must be completed.
- Sections 6 – 7: Instructions for use of EDDIES, including setting up and executing evaluation runs and analyzing the results.
- Appendices A - H: Focused and more technical material.

Two additional documents have also been developed to support EDDIES 4.4. They are likely not relevant for the average user.

- The 'Database Design' document is included in the 'EDDIES Documentation' folder of the EDDIES installation materials and describes the database tables and accounts used by EDDIES. This is intended for the advanced user who wants to query the database directly.
- A 'User Interface Specifications' document describes in detail the requirements for an EDS to be compatible with EDDIES. It is intended for EDS developers and is available upon request.

### 1.1 Online Water Quality Monitoring and Event Detection Overview

In online water quality monitoring, water quality data is collected throughout the water utility distribution system and analyzed in order to identify unusual water quality conditions. There are many potential causes of unusual water quality including cross connections, nitrification, pressure transients, water main breaks, upsets in treatment processes, and intentional introduction of foreign substances into the distribution system. If these changes are detected, utility staff can be notified to allow for investigation and timely response, if necessary.

In general, standard water quality parameters such as chlorine residual, turbidity and conductivity are monitored. These parameters have been experimentally shown to change in the presence of harmful contaminants (Hall, et al., 2007). These same parameters are useful for routine monitoring of distribution system water quality and can provide early warning of more common water quality problems.

An *event detection system* (EDS) is an automated tool for data analysis which analyzes this data in real time and generates an alert when it is deemed anomalous. Automated EDSs are broken into two major categories in EDDIES.

- **Setpoint Algorithm:** In this simple data analysis technique, an alert is produced whenever a water quality value surpasses a user-defined threshold. Alerting based on parameter setpoints/thresholds is available in most SCADA and data management systems. When deploying setpoints for event detection, the user must select the threshold values to use. A setpoint EDS is built into EDDIES: see Appendix E for more details.
- **Specialized EDS:** Several vendors and researchers have developed EDSs specifically for monitoring water quality data in real time. These use more complex mathematical and computer science approaches to time series analysis. In general, these EDSs have one or more product-specific configuration variables that impact the number and type of alerts produced. One example of an EDS configuration variable is the amount of historical data used when determining if the current data is normal or abnormal.

Identifying setpoint values or determining values for an EDS's configuration variables is called *training*. Training is generally done for each water quality monitoring location separately using historical data from that location. Depending on the EDS, training requires different levels of effort and user expertise. Some specialized EDSs “train themselves” once they are launched, whereas others require the user to do their own analyses to determine variable settings. EDDIES can be used to support training of EDSs, as is described throughout this document.

## 1.2 EDS Outputs

EDDIES collects the following outputs from EDSs. The first two outputs described are required from the EDS for each timestep; the third is optional.

- *Alert status:* a binary normal/abnormal indication for water quality. This precisely identifies when the EDS is alerting.
- *Alert level:* a real number reflecting the EDS's assessment of the likelihood that conditions are anomalous, with higher values indicating more certainty that a water quality anomaly is occurring. This measure was originally called probability of abnormality but was changed because many EDSs output values greater than one. For some EDSs (including the setpoint algorithm), the alert level is binary and is equal to the alert status.
- *Trigger parameter(s):* the water quality parameter(s) whose values caused the increased alert level. This output is optional and is generally only outputted during alerting timesteps.

In order to accurately analyze an EDS using EDDIES, the alert level and alert status must be directly related. This is true for most EDSs: an alert is produced when the alert level reaches an internal *alert threshold*. To illustrate this relationship, Figure 1-1 shows water quality data and the corresponding EDS output for a two-day period. In this example, a small drop in chlorine causes an increase in the alert level at 3/16 1:20, though the increase is not large enough to trigger an alert. However, the chlorine and TOC changes beginning at 3/16 9:00 cause an increase in the alert level large enough to trigger an alert (changing the alert status to “alerting”) at 9:55.

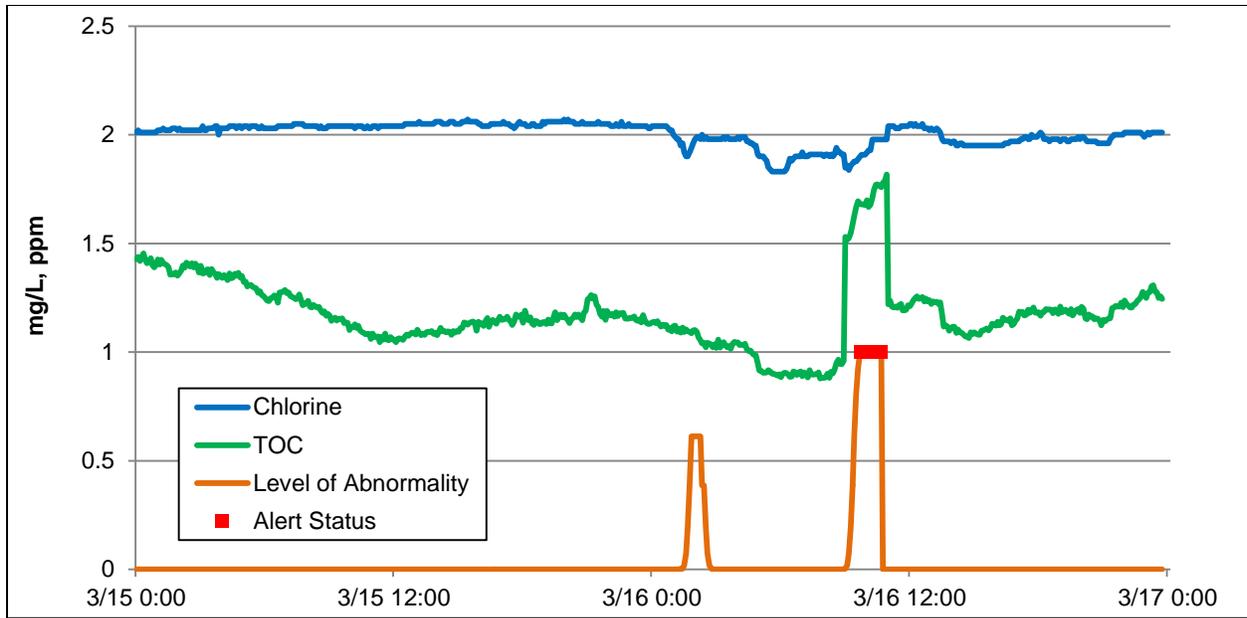


Figure 1-1. Example EDS Output during a Simulated Event

EDS output is described in more detail in Section D.1.

## Section 2.0: EDDIES 4.4 Overview

EDDIES 4.4 is an off-line tool that facilitates efficient and methodical evaluation of EDSs and EDS configurations. EDDIES allows the user to evaluate EDSs using their own data, as well as simulate contamination events on that data and capture critical performance metrics like false alarm rates and sensitivity.

Anticipated users of EDDIES include water utilities, contractors, EDS developers and researchers. Use cases of this software include:

- Comparing the performance of multiple EDSs. This can be used when selecting an EDS for purchase or implementation.
- Comparing alerting with different configurations of the same EDS. This can support the training process when implementing an EDS. It can also be useful for researchers or EDS developers in testing or refining their product.
- If using setpoint analysis for event detection, comparing alerting using different setpoint values. This can be used to select setpoint values for implementation. EDDIES contains special Setpoint Sensitivity Analysis functionality to support this, which is described in Appendix E.
- Analyzing EDS performance on various datasets. This can be used to periodically verify that the EDS configurations in use are performing as desired. For research purposes, users can consider the impact of event characteristics (e.g., contaminant type) on detection rates.
- Comparing alerting with different polling intervals. The more frequently data is collected, the more data there is to manage and store. However, longer polling intervals may result in missed detections (Umberg, 2011). This functionality may be useful for research purposes and to a utility selecting a polling interval.

A separate tool, EDDIES-RT, was developed to support real-time deployment of EDSs at water utilities. This software is not being actively supported, however, as few EDS developers have chosen to develop an EDDIES interface.

### 2.1 EDDIES Capabilities

EDDIES contains all the functionality needed to manage and implement an evaluation of EDS(s). Unique capabilities of EDDIES are described below.

#### **Evaluation using Utility Data and Simulated Events**

In order to fully characterize the performance of an EDS, both normal and abnormal datasets are needed. EDDIES allows utilities to use their own data to analyze an EDS's ability to not alert on abnormal data, as well as its ability to detect incidents of unusual water quality.

To measure an EDS's ability to detect abnormal water quality, EDDIES considers both baseline events and simulated events. Baseline events are periods the utility identifies in their own data for which an alert would be expected and desired. Examples include main breaks or treatment plant upsets. EDDIES also uses a utility's data as the basis for simulating events, superimposing water quality changes indicative of a foreign substance in the water. A detailed description of how abnormal events are simulated is given in Appendix B. The logic used is intended to produce realistic scenarios.

## **Creation and Execution of Batches**

Without EDDIES, users are generally limited in the scope of evaluation that can be performed. Manual manipulation of files is generally required for each dataset to be evaluated. The files must also be manually “run through” the EDS, and the output must be processed and analyzed individually.

With EDDIES, on the other hand, extensive evaluations can be easily implemented. The user can efficiently specify large batches of datasets to use for evaluation. If the EDS to be evaluated is compatible with EDDIES or if the EDDIES setpoint algorithm is being evaluated, the user simply clicks a button to implement the evaluation: EDDIES launches the EDS, provides all datasets to the EDS, and collects and stores the results. If the EDS is not compatible with EDDIES, one click allows the user to export the data files specified (including the simulated contamination events). They run them through their EDS externally and then with one click upload the results files into EDDIES in order to use the export and analysis capabilities described below.

## **Dataset Export and Results Analysis**

EDDIES contains several export and analysis capabilities.

- Data and Results Export: Datasets can be outputted in a variety of formats for user analysis or to run through an EDS. If the datasets have already been run through an EDS, the user can also choose to include the corresponding EDS results in these data files.
- Alerts Export: This export generates a list of all alerts produced from selected datasets, each classified as valid or invalid using the logic described in Section D.2.
- Analysis Export: EDDIES performs a variety of calculations on the EDS output from selected runs when producing this output file. Metrics include invalid alert rates, detections and detection times. Appendix D describes these analyses in detail.

## **Setpoint Algorithm and Setpoint Sensitivity Analysis**

Many utilities have the ability to define setpoints in their data management system so that an alert is generated when a specified value is surpassed. Setpoints are commonly used during typical operations to identify values outside of normal operating bounds, but these can also be used for event detection. Effectiveness of this approach depends on the specific parameter threshold values used. EDDIES allows users to evaluate a set of setpoint values. Also, the setpoint sensitivity analysis functionality allows the user to see what performance would be for a single parameter for a variety of setpoint settings. This can be useful during implementation of parameter setpoints.

## **2.2 The EDDIES Interface**

The EDDIES interface is organized in a series of menus and tabs, shown in Figure 2-1. While the use of each is described in detail throughout the document, a very general introduction is given here.



**Figure 2-1. EDDIES Tabs**

In addition to installing EDDIES (described in Section 3) and the EDS(s) to be evaluated, the user must work through the following tabs, highlighted in yellow in Figure 2-1, before evaluation can begin.

- **Location Manager Tab:** Before data can be uploaded, the user must define the parameters included in the utility data and the locations to be analyzed. A location is defined by a group of parameters that are analyzed together (which could include water quality parameters, alarm values and operational data) and generally corresponds to a particular point in the distribution system. Use of this tab is described in Section 4. Unless a user is using the Setpoint Algorithm, they likely will not use this tab after initial setup.
- **EDS Registration Tab:** EDSs to be evaluated are registered with EDDIES on this tab, described in Section 5. This tab is generally only used during initial installation and is not necessary if using the Setpoint Algorithm.
- **EDS Configuration Tab:** An EDS configuration is a specific set of values for an EDS's configuration variables. On this tab, one or more EDS configurations can be created for each registered EDS. Again, no information needs to be entered on this tab for the Setpoint Algorithm. Use of this tab is described in Section 5. A user may visit this tab often if they wish to evaluate many different configurations of EDS(s).
- **Import Manager Tab:** Data is imported into the EDDIES database using this tab, discussed in Sections 4, 5 and 6. This tab is primarily used during initial setup.

Once the data and EDSs have been defined and added to the EDDIES database, evaluation activities begin. The user creates and implements evaluations using the Batch and Launch Manager tabs and then evaluates results using the Export Manager tab.

- **Batch Manager Tab:** This is the tab where evaluation batches are created. Use of this tab is described in Section 6. This tab will likely be used often.
- **Launch Manager Tab:** This tab is used to execute the batches defined on the Batch Manager tab, either within or outside of EDDIES as described in Section 6. It will likely be used often.
- **Export Manager Tab:** Once EDS results are in the EDDIES database, this tab allows the user to export files with data, results and analysis summaries. Use of this tab is discussed in Section 7.0, and it may be used often by the user as they wish to consider EDS performance in a variety of ways.

## Section 3.0: EDDIES Installation and Configuration

Section 3.1 describes EDDIES setup and installation activities. Since EDDIES uses an Oracle database, the Oracle database must also be installed and configured before EDDIES can be used as described in Section 3.2.

The steps described in this section should be completed in the order presented. Note that EDDIES has been tested on Windows XP and Windows 7 only.

### 3.1 Oracle Database Installation and Setup

EDDIES uses Oracle's Relational Database Management System (RDBMS) as its database. If Oracle is not already installed on the workstation to be used, follow the instructions provided on the Oracle website ([www.oracle.com](http://www.oracle.com)) to install the desired version of the Oracle RDBMS (see below for details on two versions of Oracle RDBMS). It is suggested that the default installation options be used.

Attention should be paid to Oracle installation (including installation location and all passwords) as the database is difficult to modify once installed. Note that as part of Oracle installation, a password must be selected for the SYSTEM user account. It is important to remember this password, as it will be needed during the installation and use of EDDIES.

At the time of this documentation, EDDIES has been tested with Oracle 10g Express Edition, Oracle 11g Express Edition, and Oracle 11g Standard Edition One. Oracle 11g Express Edition is the most popular choice because it is available for free from Oracle's website. However, Oracle's website notes that this version is intended to be a "starter" database and will only store four gigabytes (GB) of data. This may not be enough storage for very large datasets or a large number of batches. However, the export and import scripts described in Appendix F can be used to create separate database instances, each holding only a subset of the data and results in the database at a given time and thus reducing size requirements.

Note that Oracle is upward compatible, meaning that a 10g database can be upgraded to 11g, though a database created using the 11g version of Oracle cannot be accessed via a 10g or earlier version of Oracle.

### 3.2 EDDIES Application Installation and Configuration

Once the database has been setup, EDDIES can be installed and configured. This section details the procedure for installing EDDIES and connecting EDDIES to the Oracle database.

#### 3.2.1 EDDIES Installation

It is suggested that the user close all applications running on the computer before installing EDDIES.

1. Double click the 'setup.exe' file in the EDDIES installation materials to execute the EDDIES application installation.
2. Select 'Run' on the window that pops up and then 'OK' on the next window that pops up.
3. Select a directory for the EDDIES installation from the window that opens. The default installation directory is C:\Program Files. If a different location is preferred, select the 'Change Directory' button and navigate to the desired location. Click the button with the computer icon once the desired directory is selected.
4. In the next window that opens, select a program group for EDDIES. The default installation group is EDDIES 4.4. Click 'Continue.'

The following messages are common during installation.

- If the warning message indicating the C:\WINDOWS\system32\msvcrt.dll destination file is in use pops up, select 'Ignore'. Select the 'Yes' button on the subsequent window to verify. This will not impact any computer processes and will result in a successful EDDIES installation.
- If a window pops up that says "A file being copied is older than the file currently on your system", click "Yes" to keep the existing files. Several files included in the EDDIES installation already exist on most computers.

A window will pop up indicating successful installation. Select 'OK.'

### 3.2.2 Oracle Database Setup

This section details the procedures for preparing the Oracle database for use with EDDIES. Executing the deploy.bat file creates the users and tables used by EDDIES, and the import.bat file populates default accounts and libraries (including parameter types, contaminants and event profiles). These scripts are described more in Appendix F.

#### 3.2.2.1 Deploy the EDDIES Database

The deploy script must be executed to prepare the database for use with EDDIES.

\* Note that the deploy database script deletes all data in the database tables used by EDDIES (this is only an issue if EDDIES has been used previously on the computer). However, the deploy script will not impact data in the Oracle database from other applications. Appendix F provides more details.

1. Navigate to the 'Deploy' folder within the EDDIES directory. If the default installation location was used, this will be 'C:\Program Files\EDDIES 4.4\Script and DMP File\Deploy.'
2. Execute the 'deploy.bat' batch file by double-clicking it.
3. When prompted on the DOS window that opens, shown in Figure 3-1, enter the SYSTEM account password chosen when installing the Oracle database.
4. When prompted in this DOS window, press 'Enter' to drop message objects and data that may exist from prior uses of the Oracle database. Press 'Enter' again when prompted to drop accounts and roles that may exist from prior uses of the Oracle database.
5. When prompted, choose a password for the cws\_data account. This account is not likely to be used by the average user, but the password should be noted.
6. When prompted, choose a password for the EDDIES account. This password will be used later to establish database connectivity, as discussed in Section 3.2.3.

The DOS window will automatically close when the setup is complete.

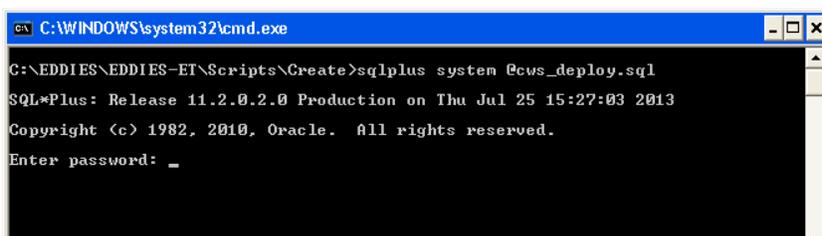


Figure 3-1. DOS Window Opened By Deploy Script

### 3.2.2.2 Import an EDDIES Database Instance

There are three options for populating the database initially.

- Import the default database DMP file included in the EDDIES installation materials, which contains the default parameter type, contaminant and event profile libraries. As this is the most common option, instructions are provided below.
- Import another EDDIES DMP file. This could have been previously generated by the user or obtained from another user. Appendix F describes the process for importing a different DMP file.
- Do not import any DMP file. Instead, start from scratch and add/import all information and libraries.

If importing the default DMP file.

1. Navigate to the 'Import' folder within the EDDIES directory. If the default installation location was used, this will be 'C:\Program Files\EDDIES 4.4\Script and DMP File\Import.'
2. Execute the import database script by double-clicking the 'import.bat' batch file.
3. When prompted on the DOS window that opens, enter the SYSTEM account password chosen during installation of the Oracle database.

The DOS window will automatically close when the import is complete.

### 3.2.3 Launching EDDIES and Connecting to the Oracle Database

EDDIES is launched by selecting a desktop icon or selecting it from the 'Start' menu. The first time EDDIES is opened, a window will pop up saying that the Oracle database is not connected. Click 'OK' and establish database connectivity as described below.

1. From the 'File' menu, select the 'Connect Oracle Database' option.
2. In the 'Database Selection' window that opens, shown in Figure 3-2, enter the following connection information.
  - **User Account:** *EDDIES*
  - **Password:** Enter the password selected in Section 3.2.2 for the EDDIES user account.
  - **Oracle Name Instance:** The Oracle name instance to be entered depends on the Oracle database version.
    - If Oracle 10g Express Edition is being used, enter *XE*.
    - If Oracle 11g Standard Edition One is being used, enter *ORCL*.
3. Select the "Connect" button. If the correct information was entered, a message will appear confirming that the connection was successful.

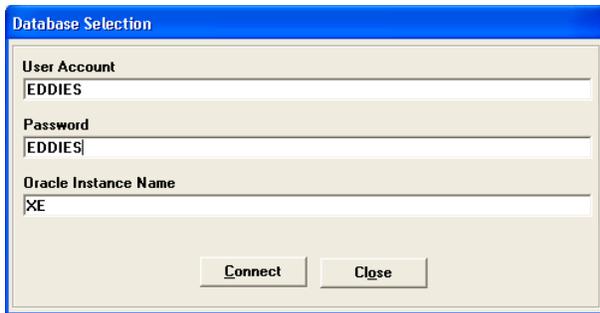


Figure 3-2. 'Database Selection' Window

## Section 4.0: Monitoring System and Utility Data Management

This section describes how to populate the EDDIES database with monitoring system information and utility data. Before EDS evaluation can begin, each of the elements described in this section must be added to the EDDIES database. They should be added in the order in which they are discussed.

\* It is highly recommended that the user review their data for periods of abnormal water quality (baseline events) prior to importing it. Once the data is imported, there is no way to change the event status to indicate baseline events. Specification of baseline events is important to accurately characterize EDS alerts: while detection of abnormal water quality is desirable, any EDS alerts during baseline events are considered false alarms instead of true detections if they are not specified by the user. Identification of periods of bad data prior to import is also important so that those alerts can also be correctly classified. Section A.5 shows how to specify the data quality in data import files and Appendix D describes how this information is used by EDDIES.

### 4.1 Monitoring System Data

Monitoring system data, listed below, defines the data streams to be used and ensures that EDDIES can accurately simulate contamination events and provide the desired data to EDSs.

- *Parameter types* describe each parameter and enable EDDIES to superimpose contamination events on the imported data. pH, tank level and sensor fault are examples of parameter types.
- *Parameters* are the individual data streams. The pH data from the West Park pump station is an example of a parameter.
- *Locations* identify the collection of data to be provided to the EDS, often corresponding to a monitoring station in the distribution system.
- *Parameter locations* map parameters to their locations. For example, the West Park pump station location may have three parameters mapped to it: that station's chlorine, pH and turbidity. Though there could be hundreds more parameters defined in the EDDIES database, datasets for that location contain only data for those three parameters.

Each subsection below details the processes of viewing, adding and editing the monitoring system data components. Data can be added using the EDDIES interface or by importing comma-separated variable (CSV) files. In general, importing data using a CSV file is more efficient for large amounts of data.

The necessary fields for each monitoring system data component are also described. For parameter types, parameters and locations, the user enters an ID and a description or name. The format of the ID field is restricted and must match the imported data, whereas the description or name field can be more descriptive. It was added because many utilities use IDs that provide little information about the parameter or location. For example, a parameter name of "Main St CL2" could be input for a parameter ID of "300008XY\_03". The name field is displayed in the EDDIES interface.

For much of the monitoring station data described in this section, changes *cannot* be made once a batch using the item has been executed. This ensures the integrity of the EDDIES database – particularly the EDS results. Changes to parameter types, parameters, locations and location parameters impact the datasets that are created, and with different datasets, the EDS outputs are different. If monitoring station data could be changed after execution of a batch using that location, the following would occur: a) there would be EDS results in the database that do not match the currently defined conditions, and b) there is no

way to determine how the system was configured for a given set of results. Thus, any analysis of EDS results is suspect.

#### 4.1.1 Parameter Types

Parameter types must be in the EDDIES database before parameters of that type can be added and, consequently, before utility data can be imported. Many common parameter types are included in the default parameter type library. Figure 4-2 shows several of these, and the full list is included in Appendix A.1. The default parameter types are also listed in the ParameterTypes.csv file located in the ‘Sample Import Files’ folder within the EDDIES directory.

The parameter types in the default library will meet most users’ needs. However, the following sections detail the how to view, add and edit parameter types if desired. The fields associated with parameter types are listed in Table 4-1.

**Table 4-1. Parameter Types Fields**

Field	Description	Required/Optional	Format
PARAMETER_TYPE_ID	Identifier associated with the data stream type to be imported.	Required	<ul style="list-style-type: none"> <li>• Text up to 100 characters</li> <li>• Cannot contain spaces</li> <li>• All letters must be capitalized</li> <li>• Must be unique</li> </ul>
PARAMETER_TYPE_DESCRIPTION	Text to be displayed in the EDDIES interface (can be the same as the ID).	Required	Text up to 100 characters
MINIMUM_VALUE	Minimum value for a parameter of the parameter type to be used during contamination event simulation.	Optional	Real number
MAXIMUM_VALUE	Maximum value for a parameter of the parameter type to be used during contamination event simulation.	Optional	Real number

The user has the option of defining a minimum and/or a maximum value for each parameter type to be used during the simulation of contamination events. If the calculated parameter value for a timestep in a contamination event simulation is less than the user-specified parameter type minimum value, the parameter type minimum value is used for that timestep rather than the calculated value. If the calculated parameter value for a timestep is greater than the user specified parameter type maximum value, the parameter type maximum value is used. If no minimum or maximum value is specified, the parameter values calculated for each timestep in an event simulation are always used. Section B.2.5 gives an example of how the minimum and maximum values are used when generating datasets.

##### 4.1.1.1 Viewing Parameter Types

Click ‘View’ from the menu toolbar. On the screen that pops up (shown in Figure 4-1), select the ‘Parameter Types’ button to view a list of the parameter types in the EDDIES database. Figure 4-2 shows the format of the view that will pop up, listing the parameter types in the EDDIES default database.

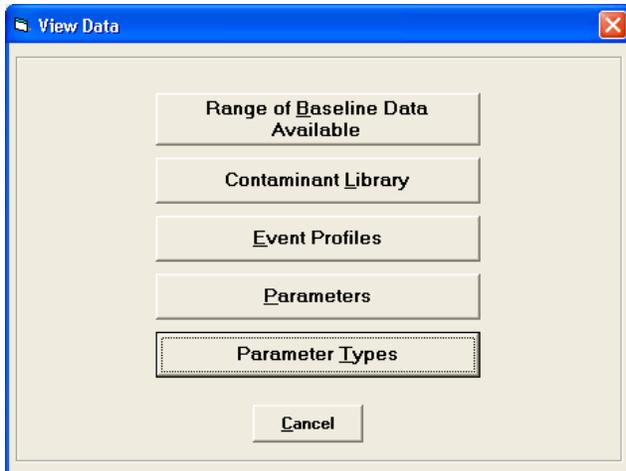


Figure 4-1. 'View' Menu

Parameter Type ID	Description	Minimum	Maximum
ALM	General Alarm	0	
ALM_BIO	Alarm produced by Biological Sensor	0	
ALM_CL-	Alarm produced by Chloride Sensor	0	
ALM_CL2	Alarm produced by Chlorine Sensor	0	
ALM_CLM	Alarm produced by Chloramine Sensor	0	
ALM_COND	Alarm produced by Conductivity Sensor	0	
ALM_DO	Alarm produced by Dissolved Oxygen Sensor	0	
ALM_FLRD	Alarm produced by Fluoride Sensor	0	
ALM_LOC	Alarm for a Particular Monitoring Station	0	
ALM_NH3	Alarm produced by Ammonia Sensor	0	
ALM_NO3	Alarm produced by Nitrate Sensor	0	
ALM_ORP	Alarm produced by ORP Sensor	0	
ALM_PH	Alarm produced by pH Sensor	0	
ALM_SEC	Security Alarm	0	
ALM_SENS	Alarm Produced by Sensor - General	0	
ALM_TEMP	Alarm produced by Temperature Sensor	0	
ALM_TOC	Alarm produced by TOC Sensor	0	
ALM_TURB	Alarm produced by Turbidity Sensor	0	
ALM_UVA	Alarm produced by UVA Sensor	0	
BIO	Biological Sensor	0	
CAL	Calibration (Normal / Calibration)	0	
CL-	Chloride	0	
CL2	Chlorine	0	
CLM	Chloramine	0	
COND	Conductivity	0	
DET_TIME	Detention Time	0	
DO	Dissolved Oxygen	0	
FLOW	Influent or Effluent Flow	0	
FLRD	Fluoride	0	

Figure 4-2. 'Parameter Types' Viewing Window

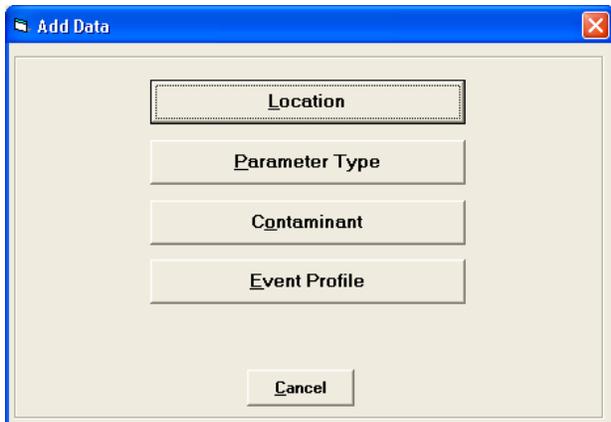
#### 4.1.1.2 Adding Parameter Types

Parameter types can be added through the EDDIES interface or using a CSV import.

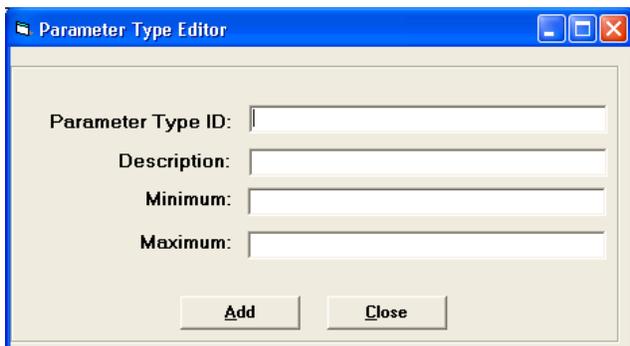
### **Adding Parameter Types using the EDDIES Interface**

Parameter types are added one at a time using this method.

1. Click 'Add' from the menu toolbar.
2. On the screen that pops up (shown in Figure 4-3), select the 'Parameter Type' button.
3. Enter the parameter type information in the 'Parameter Type Editor' window that pops up, shown in Figure 4-4. See Table 4-1 for a description of these fields.
4. Click the 'Add' button to save the parameter type information in the database.



**Figure 4-3. 'Add' Menu**



**Figure 4-4. Parameter Type Editor**

A confirmation message pops up if the parameter type is added successfully. If an error occurs, a window describing the problem pops up instead.

### **Adding Parameter Types using a CSV File Import**

Multiple parameter types can be added at once using CSV file import.

1. Create a CSV file with the desired parameter types to be added. The required file format is described in Appendix A.
2. From the 'Import Manager' tab shown in Figure 4-5, click the 'Import Parameter and Location Properties' button.
3. From the 'Import Monitoring System Data' menu that pops up, shown in Figure 4-6, select the 'Parameter Types' button.
4. Navigate to the desired CSV file in the browser window that pops up and click 'Open.'

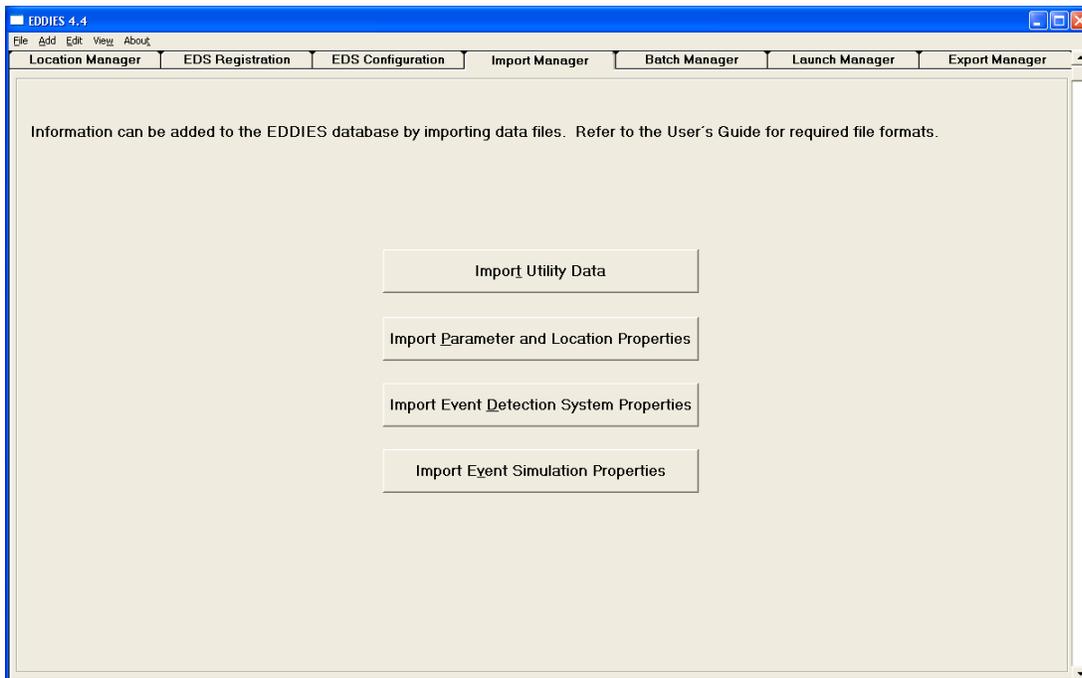


Figure 4-5. 'Import Manager' Tab

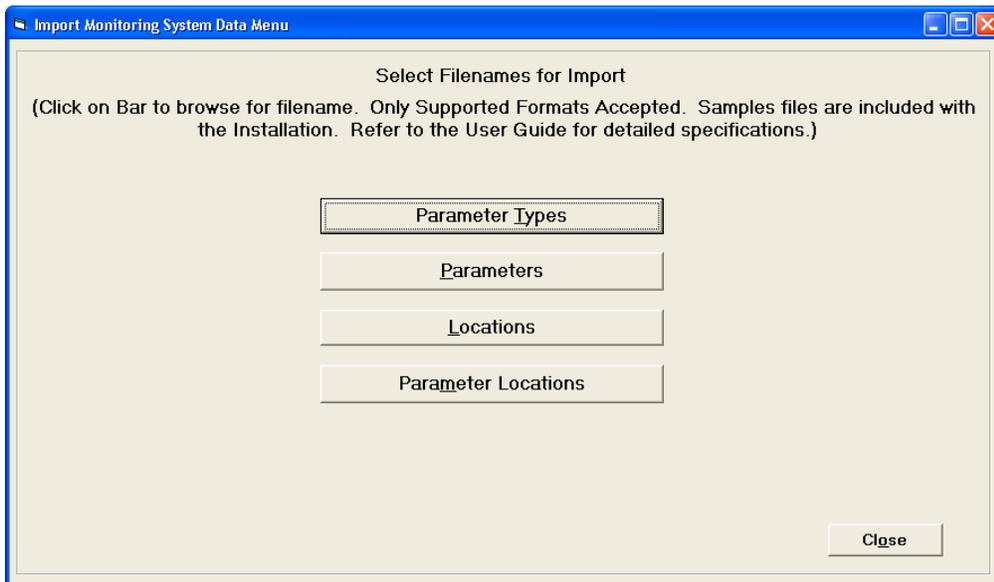


Figure 4-6. 'Import Monitoring System Data' Menu

If the import is successful, a window showing the number of parameter types added to the database appears, as shown in Figure 4-7.



**Figure 4-7. Parameter Types CSV Import Confirmation**

#### 4.1.1.3 Editing Parameter Types

Once parameter types have been added, they cannot be edited or deleted. If revisions are necessary, a new parameter type must be added.

#### 4.1.2 Parameters

Parameters must be added to the EDDIES database before utility data for those parameters can be imported. Since parameters are entirely utility-specific, no pre-defined parameters are included in the EDDIES default database.

The pertinent fields associated with parameters are listed in Table 4-2.

**Table 4-2. Parameter Fields**

Field	Description	Required/Optional	Format
PARAMETER_ID	The identifier associated with the data stream to be imported.	Required	<ul style="list-style-type: none"> <li>• Text up to 50 characters</li> <li>• Must be unique</li> <li>• Cannot contain any spaces</li> <li>• All letters must be capitalized</li> </ul>
PARAMETER_NAME	The parameter description displayed in the EDDIES Interface.	Required	Text up to 50 characters
PARAMETER_TYPE_ID	The pre-defined parameter type.	Required	Must be defined in the EDDIES database
SETPOINT_LO	The setpoint low value, used by the Setpoint Algorithm.	Optional	Real number
SETPOINT_HI	The setpoint high, used by the Setpoint Algorithm.	Optional	Real number

The Setpoint LO and Setpoint HI values are used in the Setpoint Algorithm, described in Appendix E. It is not necessary to define these values if the Setpoint Algorithm or Setpoint Sensitivity Analysis are not to be used.

The following sections detail how to view, add and edit parameters in the EDDIES database.

#### 4.1.2.1 Viewing Parameters

Parameters can be viewed using the 'View' menu or from the 'Location Manager' tab.

##### **Viewing Parameters using the 'View' Menu**

Select the 'View' menu shown in Figure 4-1 and click 'Parameters' to view a list of the parameters and their properties. Figure 4-8 shows an example of the 'Parameters' window that pops up.

Parameter ID	Parameter Name	Type	Setpoint LO	Setpoint HI
A_COND_VAL	Station A Conductivity	COND	250	500
A_CL2_VAL	Station A Free Chlorine	CL2	0.5	
A_PH_VAL	Station A pH	PH	0	
A_PRES_OP	Station A Pressure	PRES		
A_PUMP1_OP	Station A Pump 1 Status (On/Off)	PMP_STAT		
A_PUMP2_OP	Station A Pump 2 Status (On/Off)	PMP_STAT		
A_PUMP3_OP	Station A Pump 3 Status (On/Off)	PMP_STAT		
A_TEMP_VAL	Station A Temperature (Celsius)	TEMP	0	
A_TOC_VAL	Station A TOC	TOC	0	3
A_TURB_VAL	Station A Turbidity	TURB	0	5
B_CL2_VAL	Station B Free Chlorine	CL2	0.2	3.4
B_COND_VAL	Station B Conductivity	COND	200	
B_PH_VAL	Station B pH	PH	6	8.5
B_PLNT_CL2_VAL	Plant B Effluent Free Chlorine	CL2	0.3	
B_PLNT_FLOW_OP	Plant B Effluent Flow	FLOW		
B_PLNT_PH_VAL	Plant B Effluent pH	PH	0	
B_PLNT_PRES_OP	Plant B Effluent Pressure	PRES		
B_PLNT_TURB_VAL	Plant B Effluent Turbidity	TURB	0	5
B_PRES_OP	Station B Pressure	PRES		
B_TEMP_VAL	Station B Temperature (Fahrenheit)	TEMP	0	
B_TOC_VAL	Station B TOC	TOC	0	5
B_TURB_VAL	Station B Turbidity	TURB	0	
C_CL2_VAL	Station C Free Chlorine	CL2	0.2	
C_COND_VAL	Station C Conductivity	COND	284.32	396.64
C_PH_VAL	Station C pH	PH	8.12	8.64
C_PRES_OP	Station C Pressure	PRES		
C_TEMP_VAL	Station C Temperature (Celsius)	TEMP		
C_TOC_A_VAL	Station C Influent TOC	TOC		11.73
C_TOC_B_VAL	Station C Effluent TOC	TOC		
C_TURB_VAL	Station C Turbidity	TURB		

Figure 4-8. 'Parameters' Viewing Window

### Viewing Parameters from the 'Location Manager' Tab

Parameter names and types can also be viewed in EDDIES on the 'Location Manager' tab, which is shown in Figure 4-9. All parameters in the database appear in one of the listboxes (Section 4.1.4 discusses which box they appear in).

Monitoring Location ID:

(Add new locations through the menus.)

Available Parameters (with Parameter Type)

- B\_CL2\_VAL CL2
- B\_COND\_VAL COND
- B\_PH\_VAL PH
- B\_PLNT\_CL2\_VAL CL2
- B\_PLNT\_FLOW\_OP FLOW
- B\_PLNT\_PH\_VAL PH
- B\_PLNT\_PRES\_OP PRES
- B\_PLNT\_TURB\_VAL TURB
- B\_PRES\_OP PRES
- B\_TEMP\_VAL TEMP
- B\_TOC\_VAL TOC
- B\_TURB\_VAL TURB
- C\_CL2\_VAL CL2
- C\_COND\_VAL COND
- C\_PH\_VAL PH
- C\_PRES\_OP PRES
- C\_TEMP\_VAL TEMP
- C\_TOC\_A\_VAL TOC
- C\_TOC\_B\_VAL TOC
- C\_TURB\_VAL TURB
- D\_CAL CAL
- D\_CL2\_ALM ALM\_CL2
- D\_CL2\_VAL CL2
- D\_COND\_ALM ALM\_COND
- D\_COND\_VAL COND
- D\_IN1\_CL2\_VAL CL2

Selected Parameters

- A\_COND\_VAL
- A\_CL2\_VAL
- A\_PH\_VAL
- A\_PRES\_OP
- A\_PUMP1\_OP
- A\_PUMP2\_OP
- A\_PUMP3\_OP
- A\_TEMP\_VAL
- A\_TOC\_VAL
- A\_TURE\_VAL

Filter Criteria    Remove Filter    Add Parameter    Edit Parameter

Figure 4-9. 'Location Manager' Tab

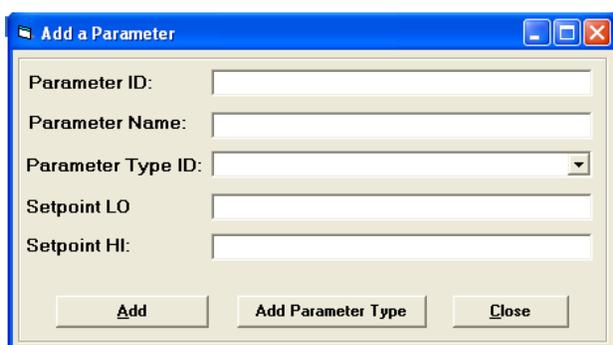
### 4.1.2.2 Adding Parameters

Parameters can be added using the EDDIES interface or through a CSV import.

#### **Adding Parameters using the EDDIES Interface**

Parameters are added to the database one at a time using the EDDIES interface.

1. At the bottom of the 'Location Manager' tab shown in Figure 4-9, click the 'Add Parameter' button.
2. In the 'Add a Parameter' window that pops up shown in Figure 4-10, enter the parameter information. Table 4-2 provides details on each parameter field. Note that the parameter type must be selected from a dropdown list. This screen includes a convenient 'Add Parameter Type' button if the desired parameter type is not yet in the EDDIES database. This button brings up the 'Add a Parameter Type' window shown in Figure 4-4.
3. Click the 'Add' button to save parameter information in the database.



**Figure 4-10. Add a Parameter**

A confirmation message appears if the parameter is added successfully.

#### **Adding Parameters using a CSV File Import**

Multiple parameters can be added at one time using a CSV file import.

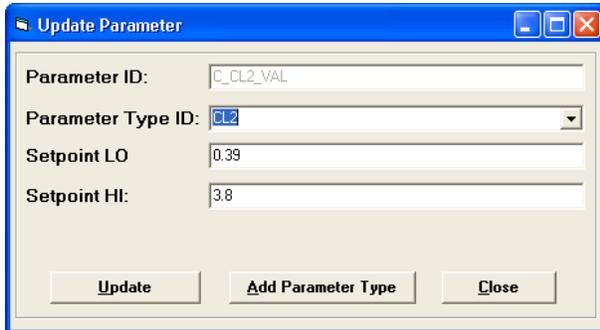
1. Create a CSV file with the desired parameters. The required file format is described in Appendix A.
2. From the 'Import Manager' tab shown in Figure 4-5, click the 'Import Parameter and Location Properties' button.
3. Select 'Parameter' from the 'Import Monitoring System Data' menu that pops up.
4. Navigate to the desired CSV file in the browser window that appears and click 'Open'.

If the import is successful, a window showing the number of parameters added to the database pops up.

### 4.1.2.3 Editing Parameters

Parameters cannot be deleted, but the parameter type and setpoints can be edited until a batch containing the parameter is completed. Once the results for a batch using a parameter are present, the fields for that parameter cannot be modified to ensure the validity of the results in the database.

1. From the 'Location Manager' tab shown in Figure 4-9, select the desired parameter in one of the listboxes so that it is highlighted in blue.
2. Click the 'Edit Parameter' button.
3. The 'Edit a Parameter' window shown in Figure 4-11 pops up with the previously defined parameter properties. Update the desired fields (the parameter ID cannot be changed). As with the 'Add a Parameter' window described in Section 4.1.2.2, parameter types can be added if needed in this window.
4. Click the 'Update' button to save the changes in the database.



**Figure 4-11. Edit a Parameter**

A confirmation message pops up if the parameter is edited successfully.

### 4.1.3 Locations

Locations must be defined in the EDDIES database before parameters can be grouped for evaluation. Since locations are utility-specific, none are included in the EDDIES default database.

The fields associated with locations are listed in Table 4-3.

**Table 4-3. Location Fields**

Field	Description	Required/Optional	Format
LOCATION_ID	The identifier associated with the location.	Required	<ul style="list-style-type: none"> <li>• Text up to 50 characters</li> <li>• Must be unique</li> <li>• Cannot contain any spaces</li> <li>• All letters must be capitalized</li> </ul>
LOCATION_NAME	The location description to be displayed in the EDDIES Interface.	Required	Text up to 100 characters

The following sections detail how to view, add and edit locations in the EDDIES database.

#### 4.1.3.1 Viewing Locations

From the 'Location Manager' tab, click the arrow to the right of the 'Monitoring Location ID' dropdown box as shown in Figure 4-12 to view a list of locations currently defined in the database.

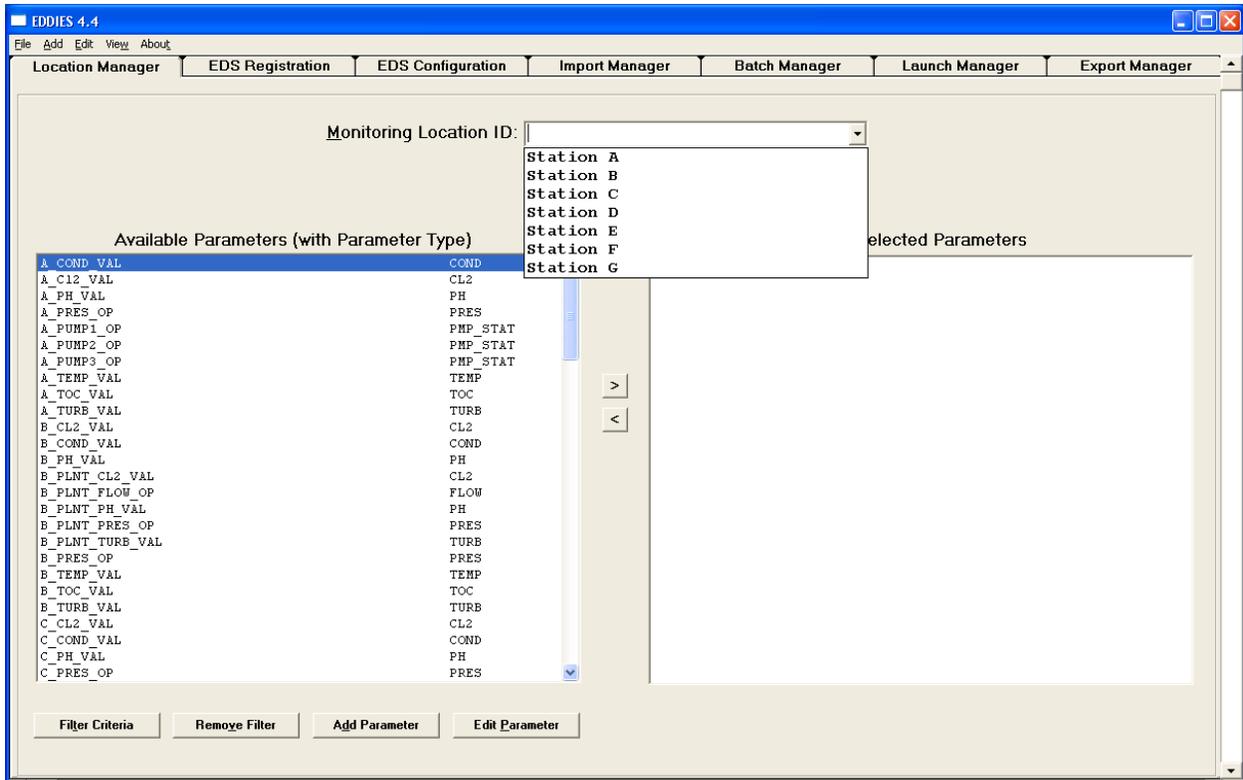


Figure 4-12. View Locations

#### 4.1.3.2 Adding Locations

Locations can be added using the EDDIES interface or through a CSV import.

#### Adding Locations using the EDDIES Interface

Locations are added to the database one at a time if the EDDIES interface is used.

1. From the 'Add' menu shown in Figure 4-3, click the 'Location' button.
2. In the 'Add a Location' window that pops up, shown in Figure 4-13, enter the location information.
3. Click 'Add' to save the location information to the database.

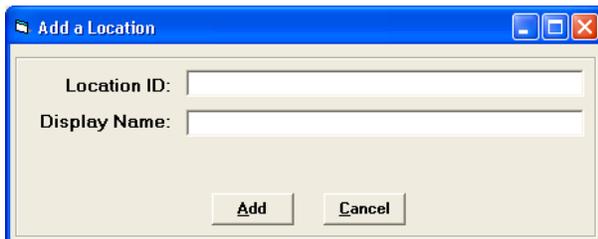


Figure 4-13. 'Add a Location' Window

A confirmation message pops up if the location is added successfully. If an error occurs, a window describing the problem pops up instead.

### **Adding Locations using a CSV File Import**

Multiple parameters can be added at one time using a CSV file import.

1. Create a CSV file with the desired locations. The required file format is described in Appendix A.
2. From the 'Import Manager' tab shown in Figure 4-5, click the 'Import Parameter and Location Properties' button.
3. Select the 'Locations' button on the 'Import Monitoring System Data' menu that pops up.
4. Navigate to the desired CSV file in the browser window that appears and click 'Open.'

If the import is successful, a window showing the number of locations added to the database pops up.

#### ***4.1.3.3 Editing Locations***

Locations cannot be edited or deleted. If revisions are desired, a new location must be added.

#### ***4.1.4 Parameter Locations***

As described in Section 4.2, each dataset contains data for the parameters mapped to the appropriate location. Note that a parameter may be assigned to more than one location. This is common with operations parameters, such as the status of a key system pump. At the same time, it is not necessary for all parameters to be assigned to a location.

The following sections detail how to view, add and edit parameter locations in the EDDIES database.

##### ***4.1.4.1 Viewing Parameter Locations***

To view the parameters currently mapped to a particular location, select the desired location from the 'Monitoring Location ID' dropdown box on the 'Location Manager' tab (as shown in Figure 4-12). The parameters assigned to the selected location will appear in the 'Selected Parameters' listbox on the right half of this tab. The parameters not assigned to the selected location remain in the 'Available Parameters' listbox on the left. Figure 4-9 shows an example where Station A is selected from the 'Monitoring Location ID' dropdown box. In that figure, there are 10 parameters mapped to Station A.

##### ***4.1.4.2 Adding Parameter Locations***

Parameter locations can be added using the EDDIES interface or through a CSV import. Unlike with parameter types, parameters and locations, the EDDIES interface is very efficient when working with parameter locations.

Before mapping can be done, all desired parameters and locations must be in the EDDIES database, as described in Sections 4.1.2.2 and 4.1.3.2.

### **Adding Parameter Locations using the EDDIES Interface**

The filter criteria functionality described below allows for efficient definition of parameter locations using the EDDIES interface.

1. On the 'Location Manager' tab, select the desired location from the 'Monitoring Location ID' dropdown box as illustrated in Figure 4-12.

2. There are two options for moving the selected parameter(s) to the ‘Selected Parameters’ listbox on the right and thus assigning them to the location. As described below, the ‘Filter Criteria’ button allows for quick search of parameters and is described below.
  - Select the desired parameter(s) from the ‘Available Parameters’ listbox (on the left) so that they are highlighted in blue. Multiple parameters can be selected at once by pressing and holding the Ctrl key or the Shift key while selecting multiple parameter names. Click the right arrow (>) button and thus assign them to the selected location.
  - Double-click on the desired parameter to move it.

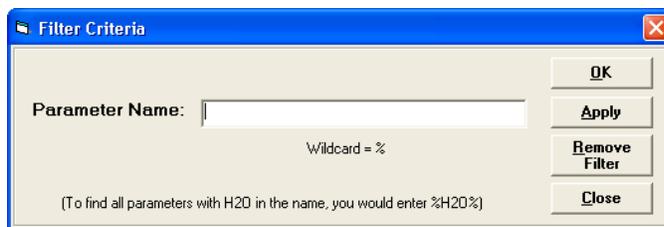
Parameters can be de-assigned from the selected location by selecting them in the ‘Selected Parameters’ listbox on the right and clicking the left arrow (<) button.

### **Filter Criteria**

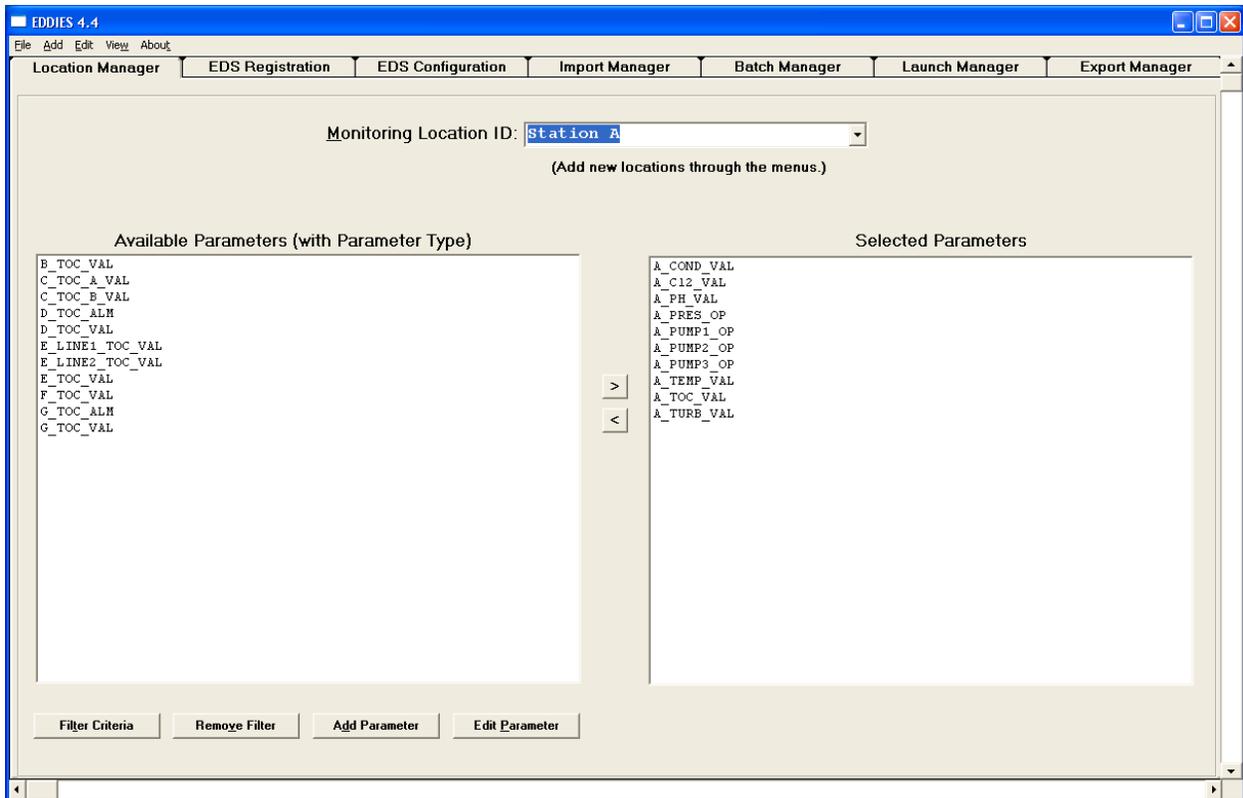
The filter criteria functionality in the EDDIES interface is used to quickly search for specific parameter names in the ‘Available Parameters’ listbox on the ‘Location Manager’ tab, shown in Figure 4-9.

Note that the filter *does not* assign the parameters to the selected location. It simply shortens the list of available parameters to make identification and selection of the desired items easier.

1. From the ‘Location Manager’ tab, click the ‘Filter Criteria’ button.
2. In the ‘Filter Criteria’ window that pops up, shown in Figure 4-14, enter search criteria for the desired parameter name criteria. Note that the percentage symbol (%) is the wildcard character. For example, %TOC% would produce all parameters with “TOC” in the parameter name. Figure 4-15 shows how the ‘Available Parameters’ listbox shown in Figure 4-9 would change if this %TOC% filter were applied. Alternately, to find parameter names that start with “A\_”, the user would enter A\_%.
3. Click the “Apply” or “OK” button to execute the filter so that only parameter IDs that meet the criteria are shown in the ‘Available Parameters’ listbox.



**Figure 4-14. Filter Criteria**



**Figure 4-15. Figure 4-9 with TOC Filter Applied**

The ‘Remove Filter’ button removes any filters that have been applied so that all parameters are shown in the ‘Available Parameters’ listbox once more.

### **Adding Parameter Locations using a CSV File Import**

Parameters can also be mapped to locations using a CSV file.

1. Create a CSV file with the desired parameter locations. The required file format is described in Appendix A.
2. From the ‘Import Manager’ tab, click the ‘Import Parameter and Location Properties’ button.
3. The ‘Import Monitoring System Data’ menu will pop up. Select the ‘Parameter Locations’ button.
4. Navigate to the desired CSV file in the browser window that appears and click ‘Open.’

If the import is successful, a window showing the number of parameter locations added to the database pops up. If an error occurs during the import, a window describing the problem pops up instead.

#### ***4.1.4.3 Editing Parameter Locations***

Parameters can be added or removed from a monitoring location until a batch containing the location is completed. The procedure described in Section 4.1.4.2 for adding parameter locations is also used to edit them.

If the user wishes to change the parameters mapped to a location, a new location must instead be created.

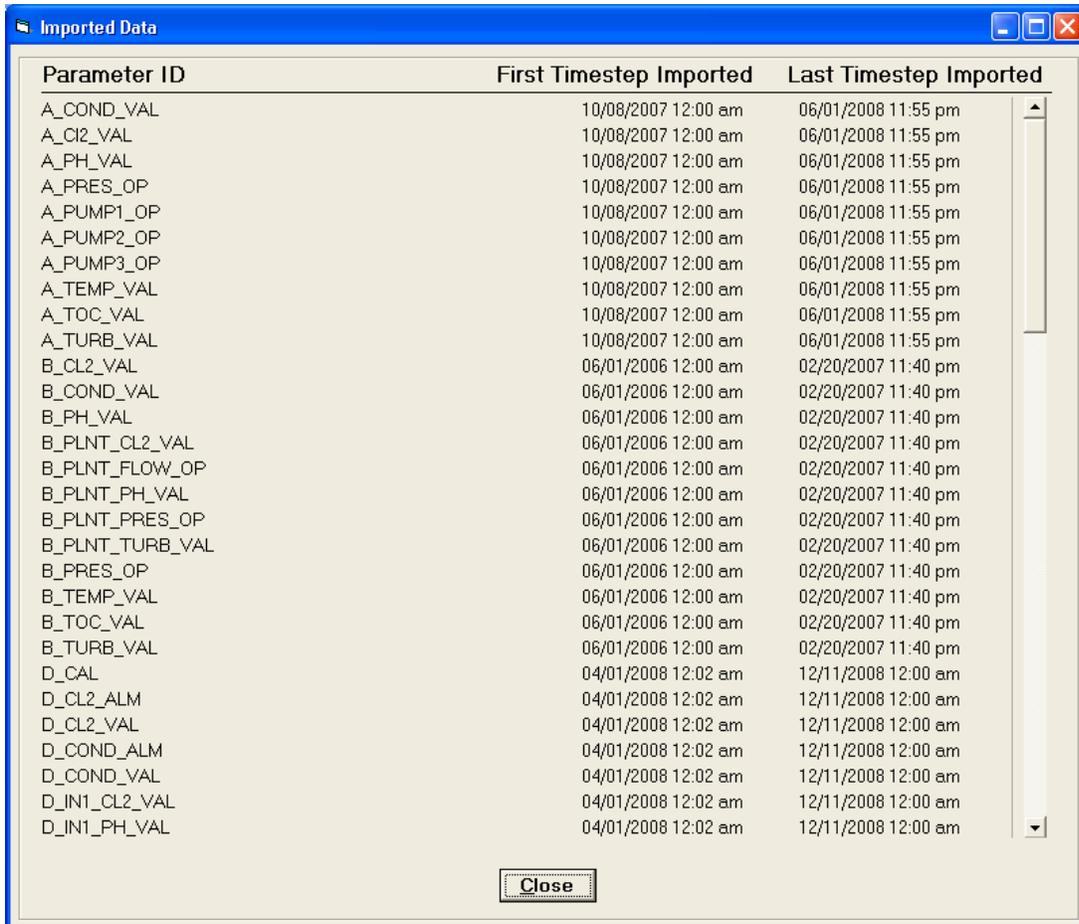
## 4.2 Imported Data

Utility data creates the basis for all evaluation datasets. Utility data is typically collected at monitoring stations and may include water quality, instrument fault and/or operations data. The fields associated with imported data are listed in Table 4-5.

### 4.2.1 Viewing Imported Data

Actual data values cannot be viewed via the EDDIES interface, as the user often uploads thousands of data points. If this detailed review is desired, a dataset export should be used (described in Section 7.0).

However, the range of dates for which there is data in the database for each parameter can be viewed. Figure 4-16 shows an example. Note that only the first and last timesteps for which there is data in the database are displayed for each parameter. There may be timesteps within this time period for which no data is stored.



Parameter ID	First Timestep Imported	Last Timestep Imported
A_COND_VAL	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_CL2_VAL	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_PH_VAL	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_PRES_OP	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_PUMP1_OP	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_PUMP2_OP	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_PUMP3_OP	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_TEMP_VAL	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_TOC_VAL	10/08/2007 12:00 am	06/01/2008 11:55 pm
A_TURB_VAL	10/08/2007 12:00 am	06/01/2008 11:55 pm
B_CL2_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_COND_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_PH_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_PLNT_CL2_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_PLNT_FLOW_OP	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_PLNT_PH_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_PLNT_PRES_OP	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_PLNT_TURB_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_PRES_OP	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_TEMP_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_TOC_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
B_TURB_VAL	06/01/2006 12:00 am	02/20/2007 11:40 pm
D_CAL	04/01/2008 12:02 am	12/11/2008 12:00 am
D_CL2_ALM	04/01/2008 12:02 am	12/11/2008 12:00 am
D_CL2_VAL	04/01/2008 12:02 am	12/11/2008 12:00 am
D_COND_ALM	04/01/2008 12:02 am	12/11/2008 12:00 am
D_COND_VAL	04/01/2008 12:02 am	12/11/2008 12:00 am
D_IN1_CL2_VAL	04/01/2008 12:02 am	12/11/2008 12:00 am
D_IN1_PH_VAL	04/01/2008 12:02 am	12/11/2008 12:00 am

Figure 4-16. View Available Imported Data

This information is accessed by going to the ‘View’ menu and clicking the ‘Range of Imported Data Available’ button.

### 4.2.2 Adding Utility Data

Utility data is added to the EDDIES database by importing CSV files. The two acceptable formats (time-series and parameter-based) are described in Appendix A. There is no way to add utility data via the EDDIES interface.

As noted in the introduction to Section 4, the event status field is important for results analysis and thus *the user should do a thorough review of the data before importing it*. Once the data is imported, there is no way to change the event status to indicate baseline events.

1. From the 'Import Manager' tab, select the 'Utility Data Import' button.
2. On the 'Import Utility Data' window that pops up (shown in Figure 4-17), specify the desired import file format. The file formats are described in Appendix A.
  - Select the 'Yes' button if the data to be imported is in time-series based format.
  - Select 'No' if the data to be imported is in parameter-based format.
3. An 'Import Format' popup message appears (as shown in Figure 4-18 or 4-19, depending on the file format selected), prompting the user to verify that the import file is in the correct format. Select 'OK' after confirming the format.
4. Navigate to the desired CSV file in the 'Select CSV File' browser window that appears.
5. Click 'Open.'

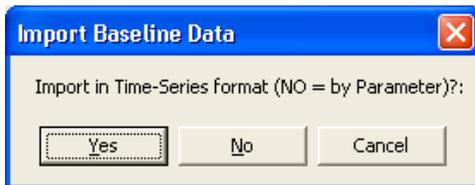


Figure 4-17. Import Utility Data (choose format)

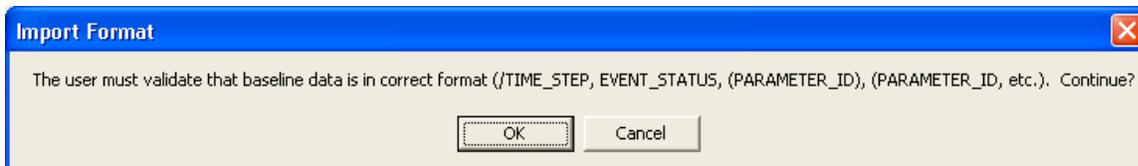


Figure 4-18. Time-Series Based Import Format Verification

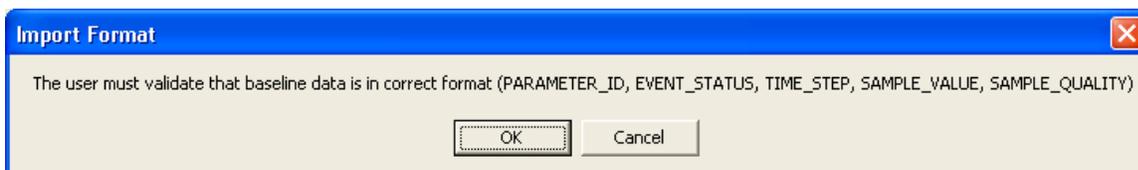


Figure 4-19. Parameter-based Import Format Verification

If the import is successful, a window showing the number of rows of data added to the database pops up. Note that import errors may occur for CSV files located on an external storage device. Move or copy the file to the C: drive to fix the problem.

### 4.2.3 Editing Imported Data

Data cannot be edited or deleted once imported.

## Section 5.0: EDS Setup

EDDIES was designed with a flexible architecture so that a variety of EDSs can be evaluated. Section 5.1 describes the three categories of EDSs that can be evaluated using EDDIES, and Section 5.2 describes the steps required to set up the EDSs for analysis.

### 5.1 Data Analysis Options

EDDIES has the capability to evaluate researched-based and commercially available EDSs. The user's guide for each EDS should be consulted to verify that the EDS can be evaluated using EDDIES and to determine if the software is EDDIES-compatible, as described in Section 5.1.1.

EDDIES also contains the capability to perform setpoint analysis. This algorithm is built into EDDIES and thus requires no setup.

#### 5.1.1 EDDIES-compatible EDS

Execution of batches using EDSs which have been made compatible with EDDIES is extremely quick and easy: no file management or interaction with the EDS is required by the user.

All setup steps described in Section 5.2 must be completed before creating and executing batches with an EDDIES-compatible EDS. Execution of batches using EDDIES-compatible EDSs is described in Section 6.2.2.1.

#### 5.1.2 External EDS

Most EDSs, however, are not directly compatible with EDDIES. As a result, this version of EDDIES was updated to allow for the evaluation of non-compatible EDSs. The EDDIES dataset creation and analysis capabilities can be leveraged, though evaluation by the EDS must be managed manually. Section 6.2.2.2 describes this process.

The EDS must meet the following requirements to be evaluated with EDDIES.

- As described in Section 1.2, the EDS's output must be directly related to degree of abnormality of the water quality data: an increased alert value must indicate a higher probability that unusual water quality is present.
- The EDS can perform offline analysis of data. Even many EDSs that are integrated into sensor hardware units have software that can be provided for use with EDDIES.
- The EDS can read files in time-series format (describe in Appendix A) and output a file containing EDS output for each timestep.
- (Ideally) The EDS will process a series of files in time-series format and produce an output file for each.

The EDS vendor should be consulted to determine the procedures for executing the EDS with datasets exported from EDDIES. The effort required by the user will vary for each EDS.

The steps in Section 5.2.3 and 5.2.4 should be followed before creating and executing batches with an external EDS. This way, the EDS name is in EDDIES and can be correctly attributed to the results.

### **5.1.3 Setpoint EDS**

Most SCADA and data management systems have the ability to generate alerts based on user-defined setpoints or thresholds. For example, setpoint alarms are often established to generate alerts if disinfectant residual levels are nearing regulatory requirements.

EDDIES includes a Setpoint Algorithm which allows the user to evaluate performance of a particular set of setpoint values. In addition, the built-in Setpoint Sensitivity Analysis Tool is intended to assist with the selection of setpoint values. A description of these capabilities and their use is included in Appendix E.

None of the steps in Section 5.2 are required for the Setpoint Algorithm or Setpoint Sensitivity Analysis.

## **5.2 EDS Installation and Registration**

This section discusses EDS installation and registration. The steps must be completed in the order listed. As noted in Section 5.1, not all steps are required for all EDSs. Each section lists the EDS type(s) for which the steps must be completed. None of these steps are required for the Setpoint Algorithm or Setpoint Sensitivity Analysis.

### **5.2.1 EDS Software Installation and Configuration**

For all EDDIES-compatible and external EDSs, the software must be installed and configured on the workstation following the instructions provided with the EDS software. Make note of the folder in which the software is installed.

Some EDSs may include specific instructions for configuration with EDDIES, including the entry of connection information for the EDDIES database. The default database configuration settings are as follows:

- Port: 1521, though a higher number may be needed if Oracle has been used on the workstation for a different application.
- Oracle Instance Name: 'XE' or 'ORCL', depending on the version of Oracle used.
- Oracle User Name and Password: EDS user account name and password chosen in Section 5.2.2.

### **5.2.2 EDS User Account**

For each EDDIES-compatible EDS, a user account must be created in the EDDIES database so it can obtain data from and post data to the database. The user should consult the EDS's documentation to see if there are any special requirements for the EDS account name and password (e.g., the CANARY EDS requires that the user name be 'CANARY'). The following IDs are prohibited because they are used by EDDIES: SYS, SYSTEM, EDDIES, CWS\_DATA and CWS\_CONNECT.

1. Navigate to the 'Deploy' folder found in the 'Script and DMP File' folder in the EDDIES program folder, chosen during installation. The default location is C:\Program Files\EDDIES 4.4.
2. Double-click the 'deploy\_eds.bat' batch file to execute it.
3. When prompted in the DOS window that pops up, enter the SYSTEM password chosen during Oracle installation (as discussed in Section 3.1).
4. When prompted, enter an account name for the EDS. The EDS account name must contain only capital letters with no spaces. It should be short and recognizable – often some form of the EDS name. It is important to save the EDS account name because it will be used by EDDIES and Oracle and it will become the EDS's Application ID within EDDIES.

- When prompted, enter a password for the EDS account. The EDS account password cannot contain spaces. It is important to save this password because it will be used by Oracle and the EDS.

The DOS window will automatically close when the EDS deployment is complete.

If account information must be modified, the steps above should be repeated.

### 5.2.3 EDS Registration

All EDDIES-compatible and external EDSs must be registered. Registration adds an EDS to the list of EDSs available for evaluation. For EDDIES-compatible EDSs, EDS registration also conveys the EDS software location to EDDIES and establishes the EDDIES/EDS connection.

The fields associated with EDS registration are listed in Table 5-1. For external EDSs, only the first field is used by EDDIES though something must be entered for all fields. The term application is synonymous with EDS.

**Table 5-1. EDS Registration Fields**

Field	Description	Example	Format
EDS (Required)	<ul style="list-style-type: none"> <li>EDDIES-compatible EDSs: the EDS account name chosen during EDS deployment (discussed in Section 5.2.1)</li> <li>External EDSs: create a new EDS identifier</li> </ul>	EDS_A	<ul style="list-style-type: none"> <li>Maximum of 50 characters</li> <li>No spaces</li> <li>All letters must be capitalized</li> <li>Must be unique</li> </ul>
App Directory (Required)	<ul style="list-style-type: none"> <li>EDDIES-compatible EDSs: the directory where the EDS software was installed</li> <li>External EDSs: any text (such as "N/A")</li> </ul>	C:\Program Files\EDS A	<ul style="list-style-type: none"> <li>Maximum of 1000 characters</li> </ul>
App Filename (Required)	<ul style="list-style-type: none"> <li>EDDIES-compatible EDSs: the name of the EDS's executable file</li> <li>External EDSs: any text (such as "N/A")</li> </ul>	EDS_A.exe	<ul style="list-style-type: none"> <li>Maximum of 1000 characters</li> </ul>
Data Directory (Required)	<ul style="list-style-type: none"> <li>EDDIES-compatible EDSs: the directory where configuration and log files should be stored. Choosing a location different than the App Directory is recommended to prevent the directory from becoming unmanageable due to the potentially large number of data files</li> <li>External EDSs: any text (such as "N/A")</li> </ul>	C:\Program Files\EDS A\Data	<ul style="list-style-type: none"> <li>Maximum of 1000 characters</li> </ul>

#### 5.2.3.1 Viewing EDS Registration Information

From the 'EDS Registration' tab, select the desired EDS name from the 'EDS' drop-down list, as shown in Figure 5-1. All previously registered EDSs will be in this list. When an EDS is selected from this list, the EDS registration information for the EDS is populated on this screen, as shown in Figure 5-2.

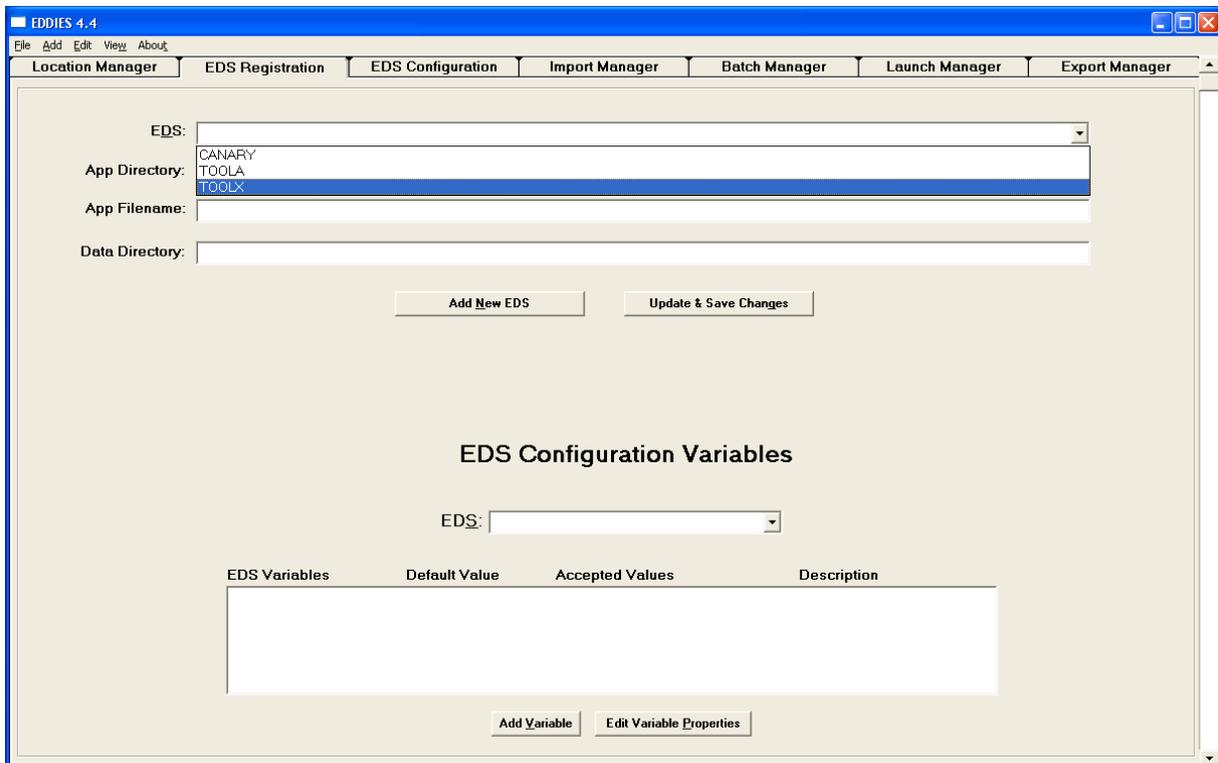


Figure 5-1. Select an EDS for Viewing

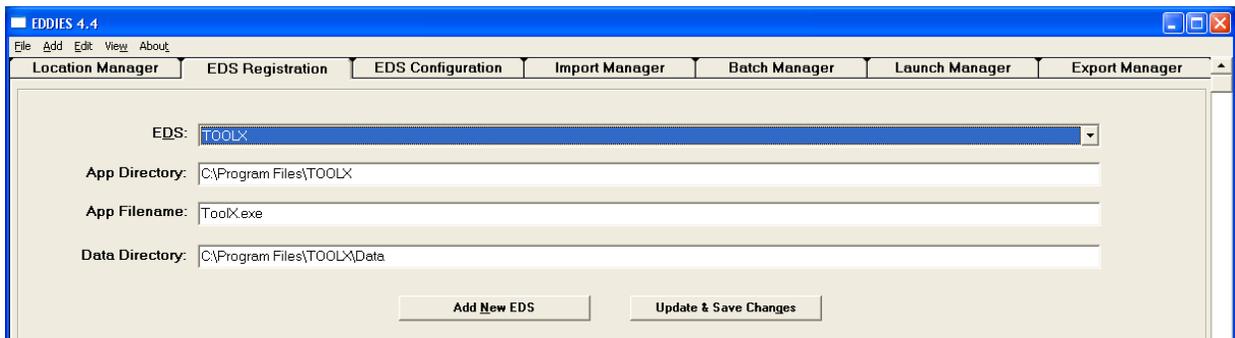


Figure 5-2. Viewing EDS Registration Information

### 5.2.3.2 Adding EDS Registration Information

EDSs can be added to the EDDIES database using the EDDIES interface or through a CSV import.

#### **Adding EDS Registration Information using the EDDIES Interface**

Registration information can be added for one EDS at a time using the EDDIES interface.

1. From the 'EDS Registration' tab, shown in Figure 5-1, click the 'Add New EDS' button.
2. In the 'Add an EDS' window that pops up, enter the EDS registration information. Table 5-1 provides details on each field, and Figure 5-2 shows example EDS information for ToolX.
3. Click the 'Add' button to save the EDS registration information in the database.

A confirmation message will appear when the EDS registration information is added successfully.

### **Adding EDS Registration Information using CSV File Import**

Two files must be imported to register an EDS using CSV file import. The required formats for these files are specified in Appendix A. EDS registration information for multiple EDSs can be added using this method.

First, an ‘EDS’ file is imported, identifying EDS ID(s) and name(s). The EDS ID is the identifier EDDIES uses internally to identify the EDS. This should be the same ID entered in Section 5.2.1. Some EDSs require a specific EDS ID: check the EDS documentation to see if there are any such restrictions. The EDS name is what the user sees on the EDDIES interface. It is intended to be more informational, and it can be the same as the ID. Next an ‘EDS Properties’ file is imported, providing EDS registration information (detailed in Table 5-2).

1. From the ‘Import Manager’ tab, click the ‘Import Event Detection System Properties’ button.
2. Select ‘EDS’ from the ‘Import EDS Data’ menu that pops up.
3. Navigate to the desired CSV file in the ‘Select CSV File’ browser that opens and click ‘Open.’ If the import is successful, a window showing the number of EDSs added to the database will pop up. If an error occurs during the import, a window describing the problem will appear instead. Any errors should be addressed before moving on.
4. Once the import is successful, click the ‘OK’ button to return to the ‘Import EDS Data’ menu.
5. From this ‘Import EDS Data’ menu, click the ‘EDS Properties’ button.
6. Navigate to the desired CSV file in the ‘Select CSV File’ browser that pops up and click ‘Open.’ If the import is successful, a window showing the number of EDS properties added pops up.

#### ***5.2.3.3 Editing EDS Registration Information***

The EDS ID and EDS name cannot be changed. The fields related to EDS execution can be updated at any time.

1. From the ‘EDS Registration tab’, select the EDS name from the EDS drop-down box as shown in Figure 5-1.
2. Update the ‘App Directory’, ‘App Filename’ and/or ‘Data Directory’ fields as desired.
3. Click the ‘Update and Save Changes’ button to save the updates to the database.

#### ***5.2.4 EDS Variables***

EDS variables must be added for all EDDIES-compatible and external EDSs. EDS variables define adjustable settings within an EDS, which tune its behavior. Performance of an EDS may vary greatly depending on the variable settings. Section 5.2.5 describes how values for these variables are set within EDDIES, but first the variables used by the EDS must be defined.

Because of how EDDIES is coded, at least one EDS variable must be defined for each EDS. For EDDIES-compatible EDSs, a list of configuration variables and guidance on populating the variable properties should be included in the EDS’s installation materials. For external EDSs, the actual values and properties are unimportant as these are not passed to the EDS: anything can be entered, including “N/A.”

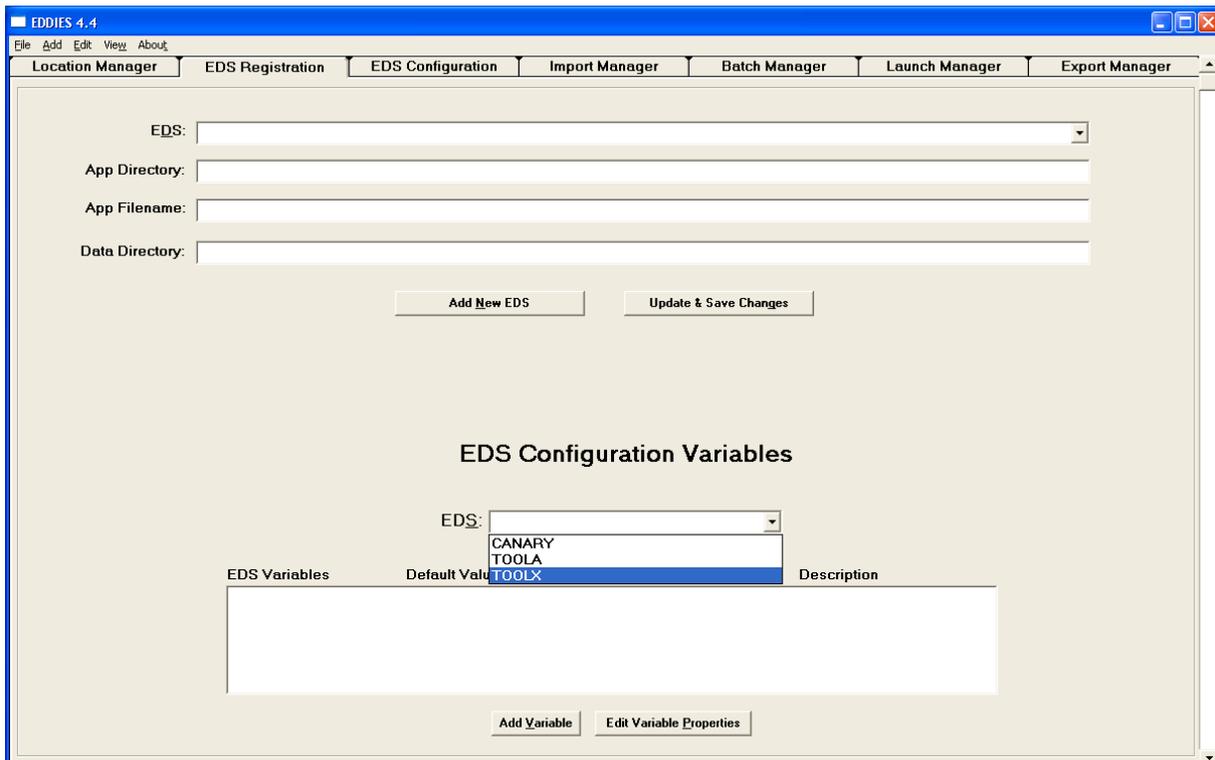
The fields associated with EDS variables are listed in Table 5-2.

**Table 5-2. EDS Variable Fields**

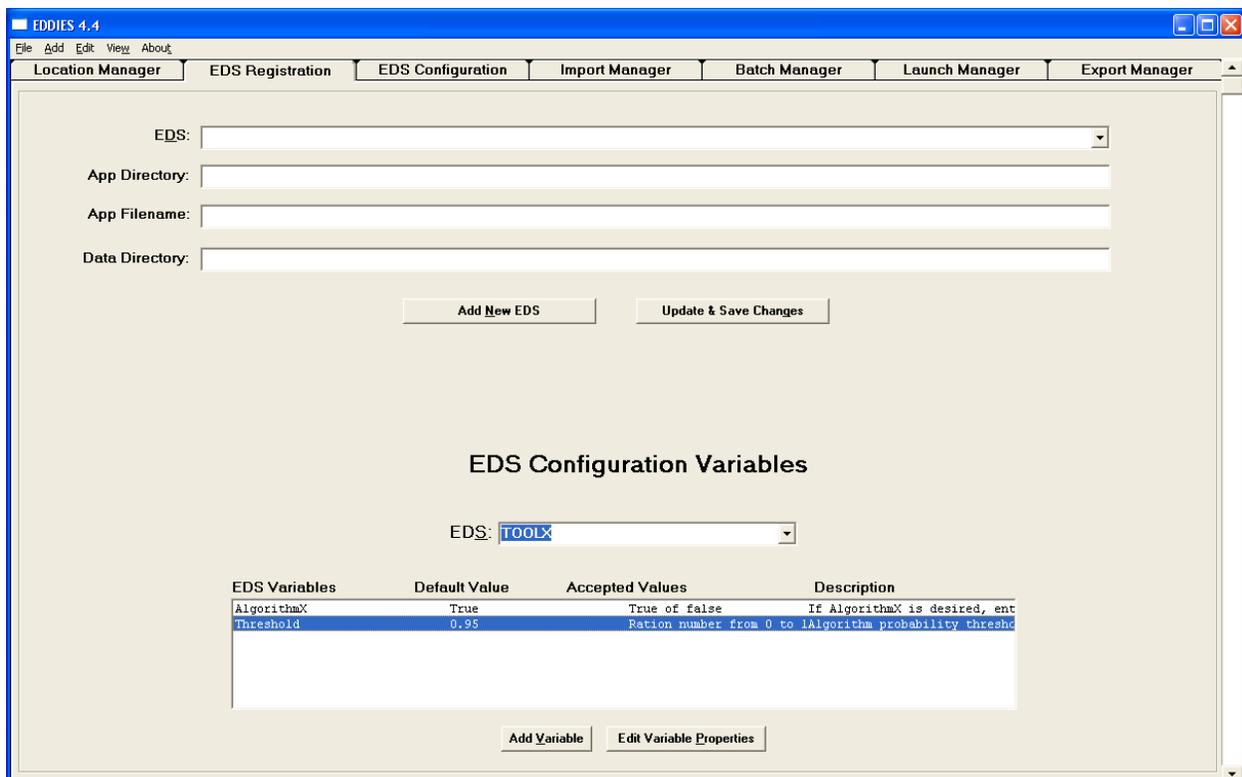
Field	Description	Example	Format
EDS (Required)	The identifier associated with the EDS. This is the identifier entered during EDS registration.	EDS_A	<ul style="list-style-type: none"> <li>Maximum of 50 characters</li> <li>No spaces</li> <li>All letters must be capitalized</li> <li>Must be unique</li> </ul>
Variable Name (Required)	The variable name that will be used by EDDIES and on all user screens. For EDDIES-compatible EDSS, specific names are often required. Check the EDS's documentation to verify.	WINDOW_SIZE	<ul style="list-style-type: none"> <li>Maximum of 1000 characters</li> </ul>
Variable Description (Optional)	A description of the variable – intended entirely for the user. This is particularly helpful when an EDS requires a more obscure variable name.	The number of timesteps of historical data considered when analyzing current data.	<ul style="list-style-type: none"> <li>Maximum of 1000 characters</li> </ul>
Acceptable Values (Optional)	A description of the acceptable values for the variable. Again, this is only for the benefit of the user.	Any integer	<ul style="list-style-type: none"> <li>Maximum of 1000 characters</li> </ul>
Default Value (Optional)	The default value assigned to the variable. If this field is left blank, the variable's value remains empty until populated by the user.	30	<ul style="list-style-type: none"> <li>Maximum of 1000 characters</li> </ul>

**5.2.4.1 View EDS Variables**

From the ‘EDS Registration’ tab, select an EDS from the drop-down box under “EDS Configuration Variables”, as shown in Figure 5-3, to view a list of EDS variables and their default settings. Figure 5-4 shows an example of the EDS variables for sample EDS, ToolX.



**Figure 5-3. Select EDS for Configuration**



**Figure 5-4. View EDS Variables**

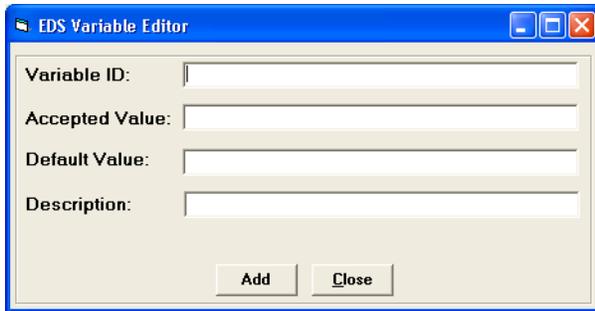
### 5.2.4.2 Add EDS Variables

This section details the process of adding EDS variables to the EDDIES database. EDS variables can be added using the EDDIES interface or through CSV file import.

#### **Adding EDS Variables using the EDDIES Interface**

EDS variables are added one at a time using the EDDIES interface.

1. From the 'EDS Registration' tab, select the name of the desired EDS in the EDS drop-down list under 'EDS Configuration Variables' as shown in Figure 5-3.
2. Click the 'Add Variable' button.
3. In the 'EDS Variable Editor' window that opens, shown in Figure 5-5, enter the EDS variable information. Table 5-2 provides details on each EDS variable field. Note that the variable name is the only required field to be entered. Click the 'Add' button to save the EDS variable in the database.



**Figure 5-5. 'EDS Variable Editor' Window**

A confirmation message will pop up if the EDS variable information is added successfully.

### **Adding EDS Variables using CSV File Import**

Multiple EDS variables for multiple EDSs can be entered at once by importing a CSV file.

1. Create a CSV file with the desired variables. The required file format is described in Appendix A.
2. From the 'Import Manager' tab, click the 'Import Event Detection System Properties' button.
3. From the 'Import EDS Data' menu that opens, select the 'EDS Configurations' button.
4. Navigate to the desired CSV file in the 'Select CSV File' browser that opens and click 'Open.'

If the import is successful, a window showing the number of EDS variables added pops up.

#### ***5.2.4.3 Edit Configuration Variables***

The EDS configuration variable name cannot be changed. However, the variable description, acceptable value, and default value can be modified at any time.

1. From the 'EDS Registration' tab, select the name of the desired EDS in the EDS drop-down list under 'EDS Configuration Variables' as shown in Figure 5-3.
2. In the listbox seen under the EDS drop-down list, highlight the desired EDS variable and click the 'Edit Variable Properties' button, or double-click the desired EDS variable.
3. In the 'EDS Variable Editor' window that opens, modify the values as desired and click 'Update.'

#### ***5.2.5 EDS Configuration***

At least one EDS configuration must be added for all EDDIES-compatible and external EDSs. An EDS configuration is comprised of a specific set of values for all of the EDS's configuration variables. For EDDIES-compatible EDSs, the user will likely create multiple EDS configurations to compare.

For EDDIES-compatible EDSs, values for each variable for a given configuration can be obtained from the EDS developer, generated by the user (the EDS software installation guide may include details on determining EDS configuration variable values), or arbitrarily chosen for testing. Most likely, at least one specific EDS configuration will be developed for each monitoring location.

For external EDSs, the user must add a configuration, though no variable values need to be entered. It is recommended that this configuration is given the same name as the EDS for clarity.

### 5.2.5.1 View EDS Configurations

EDS configurations can be viewed using the EDDIES interface.

1. From the 'EDS Configuration' tab, select the desired EDS from the 'EDS' drop-down list at the top of the screen.
2. Select the desired EDS configuration from the 'Configuration ID' drop-down list.

When the configuration is selected, the associated EDS configuration variables and their current values are shown in the listbox on this tab, as shown in Figure 5-6.

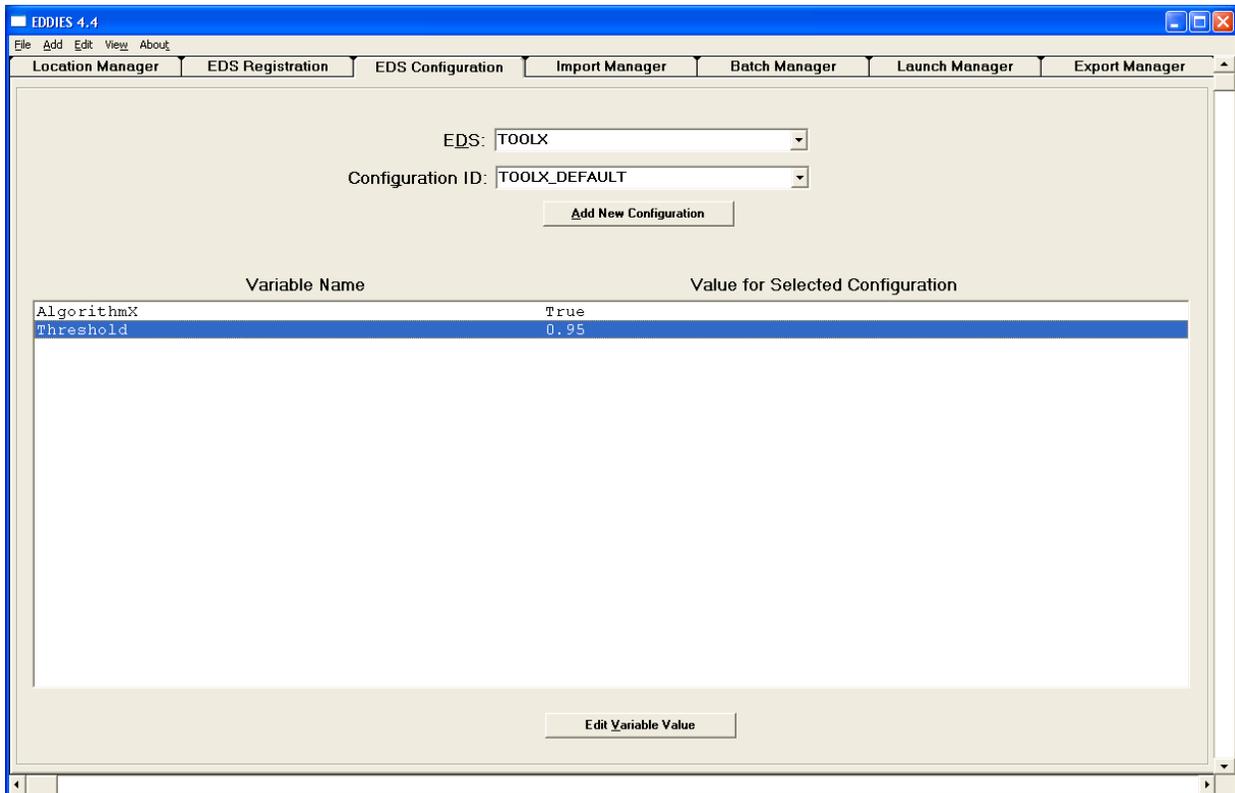


Figure 5-6. Viewing Configuration Variables for a Given EDS Configuration

### 5.2.5.2 Add EDS Configurations

EDS configurations must be added for all EDDIES-compatible EDSs and external EDSs. All EDS configuration variables defined for the selected EDS and their associated default values are included when an EDS configuration is initially created.

It is helpful to identify the EDS name, monitoring location and a description of the configuration in the configuration ID because it is used as the results identifier. Some effective examples include ToolA\_LocationX, ToolA\_DefaultValues and ToolA\_LocationX\_WindowSize30.

#### **Adding EDS Configurations using the EDDIES Interface**

Adding EDS configurations to the database using the EDDIES interface is a two-step process: first the EDS configuration ID is added and then the EDS configuration variables are defined.

1. From the 'EDS Configuration' tab, select the 'Add New Configuration' button.
2. From the 'Add New Configuration' window that opens, select the desired EDS from the 'EDS' drop-down list (all EDSs added as discussed in Section 5.3 appear in this list).
3. Enter a configuration ID and click the 'Update' button.

Once the configuration is created, the EDS configuration variable value are updated. This step is necessary only for EDDIES-compatible EDSs.

1. From the 'EDS Configuration' tab, select the desired EDS from the 'EDS' drop-down list and select the newly added configuration ID from the 'Configuration ID' drop-down list. All EDS configuration variables added as discussed in Section 5.2.4 appear in the variable listbox.
2. Select the variable to be modified in the listbox to highlight it in blue as shown in Figure 5-6.
3. Click 'Edit Variable Value' or double-click the variable name.
4. On the 'EDS Configuration Variable Editor' that opens, update the variable value as desired. The variable ID cannot be changed.
5. Click the 'Update' button.

Repeat the preceding steps for each EDS configuration variable to be modified.

### **Adding EDS Configurations using CSV File Import**

When using the CSV file import method, configurations are defined and variables are assigned values in the same step. Multiple configurations for multiple EDSs can be imported at once using this method. A new configuration will be created for each configuration listed in the import file. Note that any EDS or configuration variables included in this file must be defined before configurations can be imported.

1. Create a CSV file with the desired configurations and variable values. The required file format is described in Appendix A.
2. Click the 'Import Event Detection System Properties' button on the 'Import Manager' tab.
3. Select 'EDS Configurations' from the screen that pops up.
4. Select the desired CSV file containing the configuration information to be imported from the browser window that pops up and click 'Open'.

A window will pop up when the file has been imported successfully.

#### ***5.2.5.3 Edit EDS Configurations***

An EDS configuration – including its variable values - cannot be edited once it has been used in a batch. This ensures the integrity of the database results. If an attempt is made to modify a configuration that has already been used in a batch, an error message will appear.

If modification is desired after a batch using the configuration has been executed, a new configuration with a new name must be created. For example, assume that after running several batches with configuration ToolA\_LocationB, a new window size is to be evaluated. In this case, a new configuration with the new window size should be created (e.g., ToolA\_LocationB\_June2013Update).

However, if the EDS configuration has not been used in a batch, it can be edited on the 'EDS Configuration' tab.

1. Select the EDS associated with the configuration to be modified from the drop-down list at the top of the screen and the desired configuration from the second drop-down list. All variables and values associated with the configuration will appear in the listbox.
2. Select the variable to be modified in the listbox so that it is highlighted in blue.
3. Click 'Edit Variable Value,' or double-click the variable name.
4. On the 'EDS Configuration Variable Editor' that opens, shown in Figure 5-7, update the variable value as desired. The variable ID cannot be changed.
5. Click the 'Update' button.

## Section 6.0: Creation and Execution of Batches

Evaluation of EDSs using EDDIES is performed by creating and executing runs. A run is associated with a single dataset to be run through an EDS configuration and its corresponding results. For ease of execution, runs are grouped into batches. Each batch contains one baseline run and one or more event runs, each using the same location, time period, polling interval and EDS configuration. Batch execution is the process of sending the runs to an EDS for analysis, and then storing the results in the EDDIES database. A batch is said to be completed when all of its runs are successfully executed.

Section 6.1 describes batch creation. Section 6.2 discusses how batches are executed – both within and outside of EDDIES. This section also describes variables that impact dataset generation and batch execution.

\* It is suggested to create a database dump to back up the database before batch creation and execution. Saving the fully configured database with monitoring system information and uploaded data before the EDS results are imported provides insurance for the possibility of a database corruption. In addition, multiple databases for different EDSs or evaluations can be created. See Appendix F for more details.

Note that only one batch creation and/or execution session can take place on a single computer at a time.

### 6.1 Create a Batch of Runs

Batches are created through the EDDIES interface using the ‘Batch Manager’ tab or by importing a CSV file containing batch information. Appendix B describes the run properties and how events are simulated. Runs within batches are automatically assigned Run IDs as sequential integers, beginning with the baseline run.

#### 6.1.1 Batch Properties

The following batch properties are populated to create a batch:

- **Batch ID:** an alphanumeric string with no spaces, used to identify the batch. It must be unique and it is helpful if this ID is descriptive (e.g., ToolA\_LocationA\_LargeEvents).
- **Location:** the location to be analyzed, which determines the parameters to be provided to the EDS for analysis. Any location previously added (as described in Section 4.1.3) can be chosen.
- **Polling Interval:** any positive integer defining the timestep frequency (in minutes) at which the water quality data is reported in the datasets for EDS analysis and for dataset export. Section B.1 further describes how this batch characteristic impacts the datasets.

To enable a more accurate evaluation, the polling interval should be set no less than the frequency at which the key parameters are reported in the utility data (e.g., if the utility data has data every five minutes, do not use a polling interval of less than five minutes). If a set polling interval is already used at the utility, that should be chosen as the polling interval for the batch unless an analysis of the impact of changing the polling interval is desired.

- **Configuration ID:** the EDS configuration used to analyze the batch datasets. Any EDS configuration previously created (as described in Section 5.2) or the Setpoint Algorithm or Setpoint Sensitivity Analysis Tool (described in Appendix E) can be selected. A configuration ID must be selected even if the batch is to be executed outside of EDDIES. Note that the EDS configuration chosen inherently determines the EDS to be used.

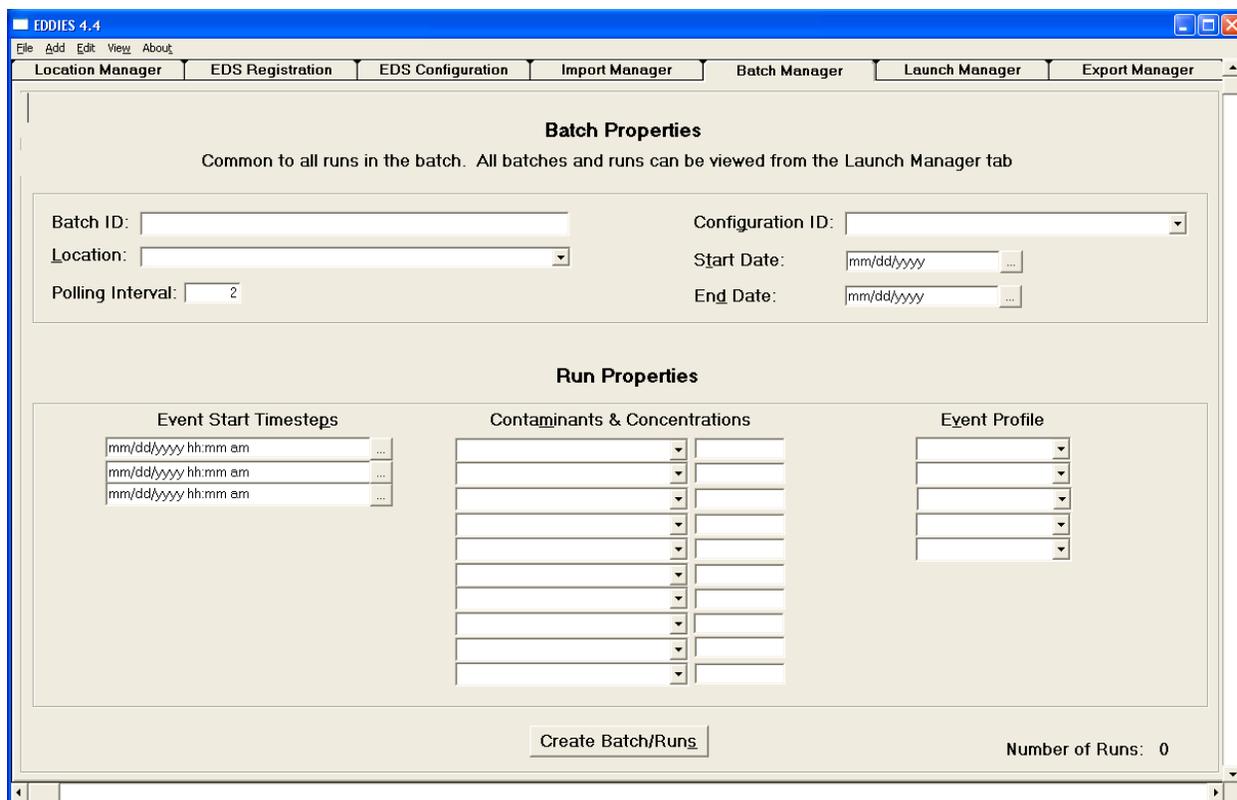
- **Start Date and End Date:** the date range for the baseline dataset to be presented to the EDS, inputted in the format MM/DD/YYYY. All event start times must also occur within this time period. Baseline datasets start at 12:00 AM on the start date and end on the last timestep of the end date according to the polling interval. If the specified start or end date is outside of the range of baseline data available in the EDDIES database, an error message will be generated. Section 4.2 describes how to review the baseline data in the database.
- **Run Properties:** variables that define the simulated contamination scenario(s) for the batch, consisting of the characteristics below. At least one simulated event must be defined for each batch.
  - **Event Start Timesteps:** timestep(s) in the baseline data at which event simulation begins, input in the format MM/DD/YYYY hh:mm am (12-hour format). This timestep must fall between the batch's start and end dates. If a start time is entered that does not match a baseline dataset timestep generated based on the specified polling interval, EDDIES adjusts the event start time to the first relevant timestep after the event start time entered. For example, if an event start time of 2:03 AM is entered for a batch with a five minute polling interval, EDDIES adjusts the event start time to 2:05 AM.
  - **Contaminants:** contaminant(s) to be used in event simulation, selected from a dropdown list of all the contaminants defined in the Contaminant Library.
  - **Concentrations:** peak contaminant concentration(s) at which events are to be simulated. These can be any non-negative number(s).
  - **Event Profiles:** event profile(s) to be used in event simulation, selected from a dropdown list of all profiles defined in the Event Profiles Library.

## **6.1.2 Methods for Defining a Batch**

A batch can be created using the EDDIES interface or through a CSV import, as described below. Note that the number of runs significantly impacts the time required to execute the batch.

### **6.1.2.1 EDDIES Interface**

Batches can be efficiently created using the 'Batch Manager' tab, shown in Figure 6-1. The user enters the desired batch properties and a run is created for *all combinations* of the entered run properties. Thus, hundreds of runs can be quickly generated.



**Figure 6-1. Batch Manager**

First, enter the batch properties, the batch ID, location, polling interval, configuration ID, start date and end date at the top of the ‘Batch Manager tab’. These selections apply to all runs in the batch.

Next, define the run properties, consisting of event start timestep, contaminant, contaminant concentration and event profile. At least one value must be defined for each run property, and EDDIES creates a run for every combination of the entered event start time, contaminant/concentration pair and event profile. The same contaminant may be entered multiple times with different concentrations.

The counter on the bottom right of the tab displays how many runs will be created based on the entered properties. A maximum of three event start timesteps, 10 contaminant/concentration combinations and five event profiles may be entered using the EDDIES interface. If more run properties are desired for a single batch, the CSV file import batch creation method should be used.

Once all desired entries are made, click ‘Create Batch/Runs’. A message will appear when the batch is successfully created. No runs can be added to the batch after the batch is created.

### 6.1.2.2 CSV File Import

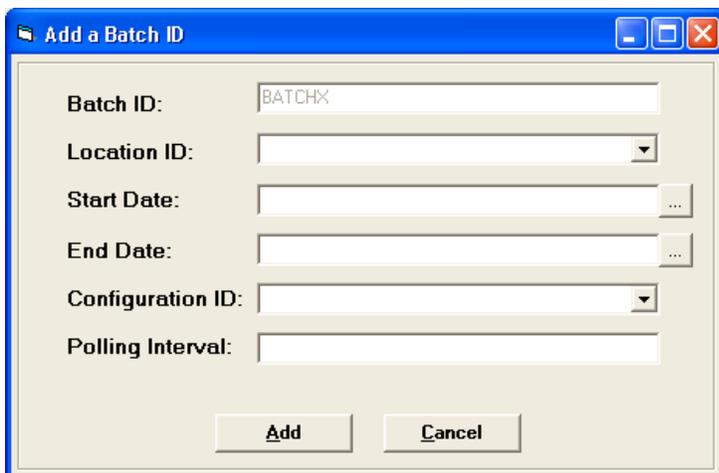
Multiple batches can be created with a single CSV file, and EDDIES creates a run for each row in the import file. It is not necessary for a batch created through CSV import to consist of all combinations of the entered run properties.

1. Create a CSV file with the desired runs. The required file format is described in Appendix A.
2. Select ‘Import Event Simulation Properties’ from the ‘Import Manager’ tab to obtain the ‘Import Analysis Data’ menu shown in Figure 6-2.



**Figure 6-2. Analysis Data Import Menu**

3. Select 'Batch/Runs'.
4. Locate the desired file to import using the 'Select CSV File' browser that pops up.
5. For each Batch ID included in the file, a pop up window will appear (shown in Figure 6-3). Enter the overall batch properties (described in Section 6.1) and click 'Add.'



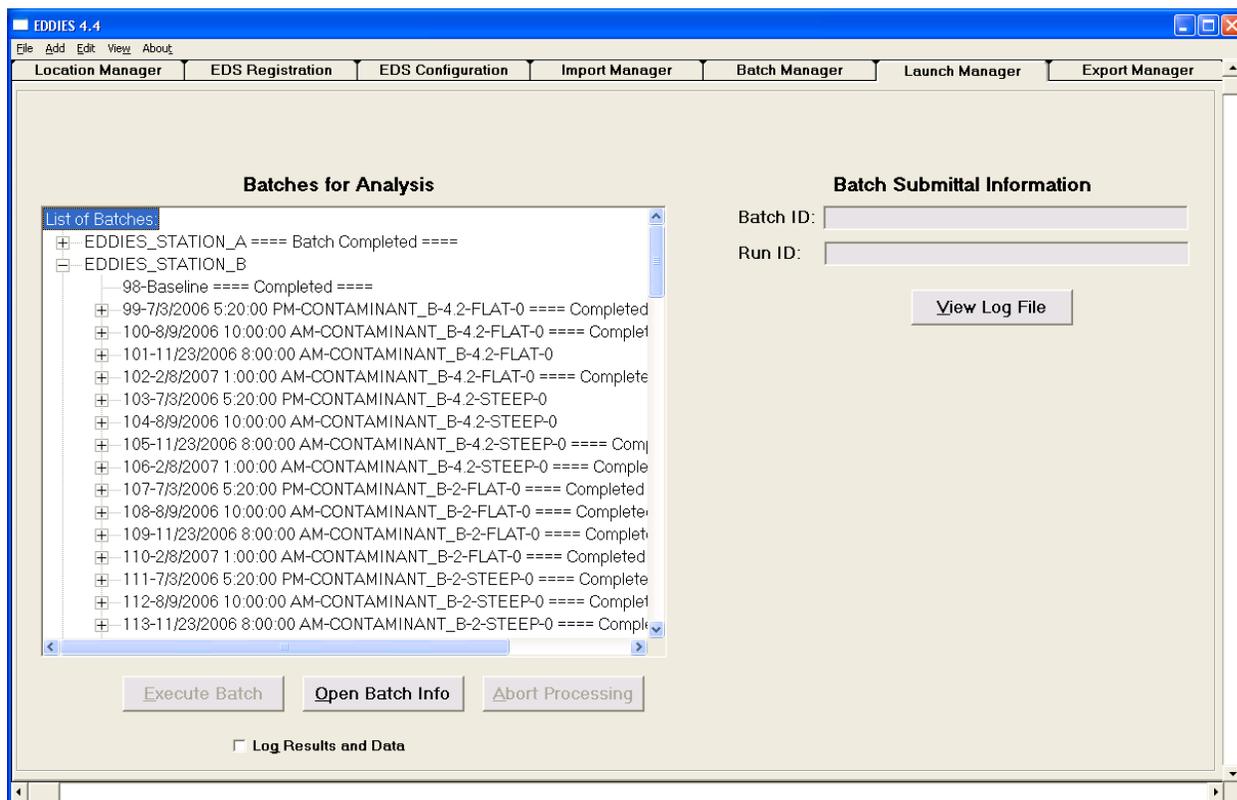
**Figure 6-3. Enter Batch Information**

Once the batches are added, a message will appear confirming that the batch is successfully created and displaying the number of runs within the batch.

## 6.2 Execute a Batch

Once a batch is created, it appears on the 'Launch Manager' tab as shown in Figure 6-4. The 'List of Batch IDs' presents an outline view of the created batches. Clicking the '+' sign next to each Batch ID

expands the item to reveal a description of each run in the batch. Figure 6-4 shows an example of the ‘Launch Manager’ tab where the “EDDIES\_STATION\_B” batch is expanded.



**Figure 6-4. Launch Manager**

The first run listed in a batch is always the baseline run, simply described as “Baseline” in the outline view on the ‘Launch Manager’ tab. The event runs follow sequentially, and their descriptions are concatenations of the run characteristics separated by dashes.

Batches for which results exist for all runs in the EDDIES database are shown by the adjacent indicator, ‘===== Batch Completed =====’, as demonstrated by the “EDDIES\_STATION\_A” batch in Figure 6-4. For partially completed batches, ‘=====Completed =====’ is shown next to the completed runs when the batch is expanded.

### 6.2.1 Batch Variables

Before executing a batch, the values of the variables that impact dataset creation (Section 6.2.1.1) and batch execution (Section 6.2.1.2) should be verified.

#### 6.2.1.1 Event Simulation Variables

For event runs, EDDIES only considers EDS results for event simulation timesteps, as described in Appendix D. The Start and End EDS Analysis variables allows the user to define a time period around the contamination scenario so that only a subset of the baseline dataset is generated for each event run – providing the EDS enough data to ensure valid output while eliminating unnecessary processing. This can significantly reduce batch processing time.

- **Start EDS Analysis:** This variable defines the number of days of the baseline dataset prior to the simulated contamination event for each event dataset. The event dataset for each run always

begins in a new day. For example, if the Event Start Time for a specific run is set to 2/12/2009 1:21 PM and the Start EDS Analysis duration is set to one day, the event dataset would begin on 2/11/2009 at 12:00 AM. If the Start EDS Analysis variable is left blank, the entire duration of baseline dataset is provided for each event run.

Many EDSs require an initialization period in order to produce meaningful results. For example, an EDS may need several hours or days of data to generate baseline statistics for use in its subsequent analysis. Therefore, *the Start EDS Analysis variable must be chosen carefully* to ensure accurate results. Refer to the EDS's documentation for guidance on selecting this value.

- **End EDS Analysis:** This variable defines the number of days of baseline data after the simulated contamination event ends for each event run. The event dataset for each run always ends on the last timestep of the day according to the batch polling interval. Consider an example run whose contamination simulation scenario ends on 5/1/2011 9:20 AM in a batch assigned a five minute polling interval. If the End EDS Analysis duration is set to one day, the event dataset would end on 5/2/2011 at 11:55 PM. The default is zero, meaning that the run ends on the day in which the contamination event ends, as all necessary output is obtained from the EDS at that point.

To view and edit these variables, select 'Event Simulation Variables' from the 'Edit' menu, shown in Figure 6-5. Enter the desired values in the 'Analysis Settings' window that pops up and select 'Update'.

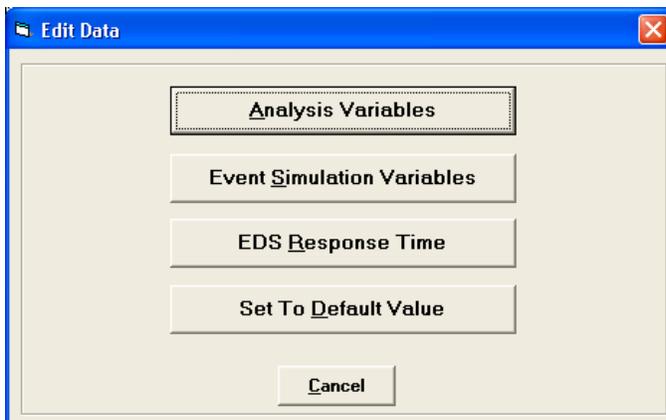


Figure 6-5. 'Edit' Menu

### 6.2.1.2 EDS Response Time Variable

The EDS response time variable is relevant only when using an EDDIES-compatible EDS. EDDIES directly communicates with the EDS when a batch is executed in this way. The EDS response time seeks to avoid a situation where the EDS locks up or shuts down while EDDIES indefinitely remains unaware of the problem, waiting for a response. This variable in no way affects the datasets or results.

The EDS response time is the maximum number of minutes the EDS has to respond after EDDIES sends a message. If the EDS does not send a response within this amount of time, EDDIES aborts the run and an error message pops up to indicate that the batch failed and to ask if the results associated with that run should be deleted.

To view or edit the response time, select 'EDS Response Time' from the 'Edit' menu. Enter the desired value in the window that pops up and select 'OK'.

Suitable EDS response time values vary based on the number of timesteps in each run and the speed with which the EDS processes the data. The response time should be increased if multiple batches are aborted by EDDIES because the EDS response time was exceeded.

### **6.2.2 Executing a Batch**

Batches can be executed within EDDIES or the EDS analysis can be performed externally and the results imported to the EDDIES database. Executing a batch through EDDIES is less effort for the user but requires EDS compatibility with EDDIES.

A batch can be re-submitted after it has been completed, though the results for completed runs within that batch are deleted. If the user chooses to execute a batch for which results exist in the EDDIES database, a window will pop up notifying the user that all results for the batch will be deleted prior to re-execution. However, there are checks within EDDIES to ensure that information affecting batch results is not modified once a batch or run is completed. Therefore, there is no benefit in re-running a batch since all components of the batch, and thus the results, would be the same. If the user would like to run a batch with slightly different settings, a new batch containing the updated settings should be created.

#### **6.2.2.1 Executing a Batch within EDDIES**

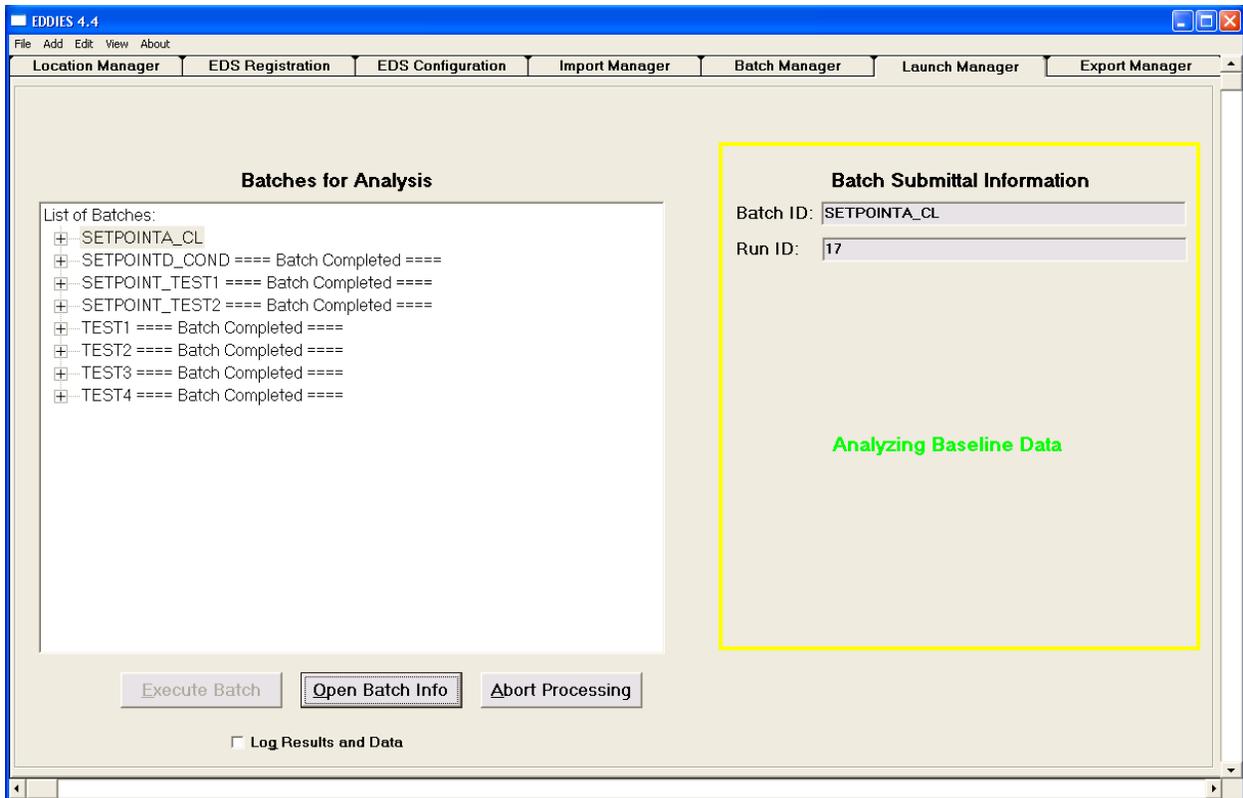
Batches are executed through EDDIES via the ‘Launch Manager’ tab, shown in Figure 6-4. Once a batch is submitted within EDDIES, EDDIES completes the following processes and no additional action is required of the user.

- Starts the EDS and provides the EDS with the desired configuration settings.
- Generates the dataset for the first uncompleted run and provides it to the EDS.
- Collects and stores the EDS output.
- Repeats the previous two steps for all runs in the batch.
- Shuts down the EDS when results from all runs are successfully stored.

Before submitting a batch, the user can choose to click the ‘Log Results & Sample Data’ checkbox at the bottom of the ‘Launch Manager’ tab. Doing so opens the ‘Select Output Folder’ browser where the user selects a directory in which to save CSV files containing results and data for each run. As the batch is executed, one CSV file is written to the selected directory for each run; the CSV file contains the dataset EDDIES presented to the EDS for that run with the corresponding EDS output. The file name is the numeric Run ID. Appendix C outlines the format of this CSV file.

To execute a batch, select the desired batch from the listbox on the ‘Launch Manager’ tab so that the batch name is highlighted in blue and select ‘Submit Batch’. A message appears noting the batch execution start time. Click ‘OK’ to verify batch execution initiation. Only one batch can be submitted and/or running at a workstation at a time.

After the batch is submitted, the progress of the batch is shown on the right side of the ‘Launch Manager’ tab, as shown in the boxed portion of Figure 6-6. If any errors are encountered, error messages are displayed in red in the same area. The batch and run in process is displayed in ‘Batch ID’ and numeric ‘Run ID’ metrics, respectively, allowing the user to track the progress of the batch execution.



**Figure 6-6. Batch Processing Progress**

When all runs in the batch are successfully completed and stored in the database, a popup window appears confirming that the batch is complete.

Clicking the ‘Abort Processing’ at the bottom of ‘Launch Manager’ tab at any time will stop batch execution and shut down the EDS. When this is done, the user is asked whether the results should be deleted.

- If ‘Yes’ is selected, all results from the batch are deleted. If the user later re-submits the batch, EDDIES processes the entire batch, beginning with the baseline data. The user is able to modify the components of the batch, such as the parameters mapped to the batch’s location, as long as they have not been used in a completed batch.
- If ‘No’ is selected, results from fully completed runs are saved. If the user subsequently re-submits the batch, EDDIES begins execution with the first incomplete run. This option is useful due to the lengthy amount of time necessary to run a batch: if the EDS locks up or EDDIES shuts down in the middle of a batch execution, the entire batch does not need to be re-run. Data and results from completed runs can be exported (as described in Section 7).

### 6.2.2.2 Executing a Batch Outside of EDDIES

Batches can also be executed outside of EDDIES, allowing the user to evaluate EDSs that are not compatible with EDDIES while leveraging EDDIES’s dataset generation and results analysis functionalities.

It is up to the user to ensure the objectivity of a batch run outside of EDDIES, particularly if the analysis is not executed by the user. Simulated events are often easily identified by plotting the water quality data, and thus it is feasible that the EDS output files could be manipulated to improve performance. This is especially relevant if different EDSs are being compared.

To execute a batch outside of EDDIES, the user creates a batch as described in Section 6.1. All components of a batch, including EDS Configuration information, are necessary. Since the EDS is to be executed outside of EDDIES and therefore does not require communication with EDDIES, the EDS registration and configuration information entered, such as app filename and directory, is arbitrary.

The following steps describe how to execute a batch outside of EDDIES. These steps are divided into three stages: exporting test datasets, running the data through the EDS, and obtaining and importing results. Verify the values of the Event Simulation Variables (described in Section 6.2.1) before beginning an export.

#### 1. Exporting Test Datasets

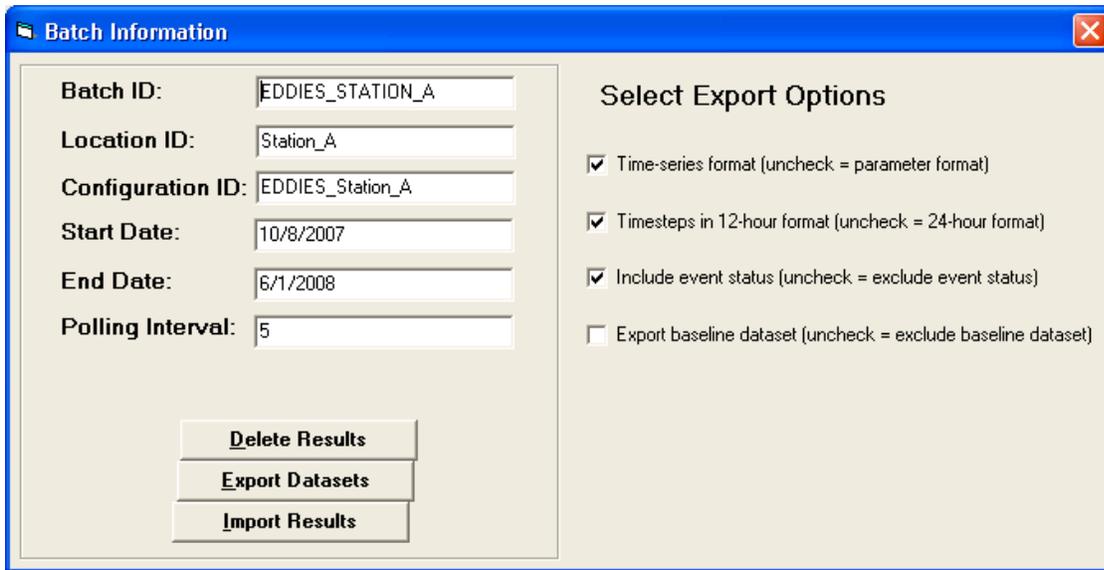
- a. On the ‘Launch Manager’ tab, select the desired Batch ID in the listbox so that it is highlighted in blue. Right-click the Batch ID or click the ‘Open Batch Info’ button.
- b. The ‘Batch Information’ window (shown in Figure 6-7) will appear. Select the desired export options using the check boxes on the right side of this window:
  - File format: The files can be exported in time-series or parameter-based format. Examples of these formats are shown in Appendix C.
  - Timestep format: The timesteps can be provided in 12-hour (AM/PM) format or 24-hour (military time) format.
  - Event status: Event status can be included as the final column. This should not be selected if the data files are to be analyzed by an EDS.
  - Baseline data: A file containing the baseline data can be generated or excluded. If this box is not checked, only event datasets are exported. This option was added because baseline datasets are often large and can take a long time to generate.
- c. Select the ‘Export Test Datasets’ button.
- d. Select the folder to which the data files should be written in the browser window that pops up. One file is outputted to this folder for each run of the batch. The files are automatically named as ‘Run\_X\_Dataset’ where X is the Run ID. For example, the file name for Run 4 would be ‘Run\_4\_Dataset’.

#### 2. Processing the Data and Obtaining Results

- a. Run the exported dataset files through the desired EDS. Consult the EDS documentation for details on how this is done.
- b. For each dataset, the EDS should produce a file containing corresponding results named ‘Run\_X\_Results’ where X is the Run ID. Save the files of all batch runs in a single directory, but not in the directory in which the batch datasets are located (‘Run\_X\_Results’ should not be in the same folder as ‘Run\_X\_Dataset’). See Appendix A for guidelines on the required results import file format. The exported results files may need to be updated if the EDS output files do not match the results input format required by EDDIES.

### 3. Importing Results Files

- a. Return to the 'Launch Manager' tab and once again select the desired batch so it is highlighted in blue. Right-click the batch ID or click the 'Batch Information' button.
- b. Select 'Import Results' on the 'Batch Information' window (shown in Figure 6-7) that pops up.
- c. Navigate to the directory where the EDS results files are located in the browser window that pops up and select 'OK'.



**Figure 6-7. Batch Info**

EDDIES provides a completion indication when all results are successfully imported in the form of the red "DONE" message below the batch information on the 'Batch Information' window.

Results files for a single batch can be imported in stages: results files for any number of runs within a batch can be imported initially and results files for the remaining runs can be imported at a later time. If the directory selected by the user for results import contains results file(s) for run(s) that have previously been saved in the database, EDDIES asks the user if such results should be skipped or re-imported.

- If the user chooses to skip the file(s), the existing results remain in the database and EDDIES goes on to the next results file. This option is suggested unless the output in the files have been updated.
- If the user chooses to re-import the file(s), the previous results are deleted and the results file is imported.

## Section 7.0: Export and Analysis of Results

Data in the EDDIES database can be exported for viewing and analysis through the ‘Export Manager’ tab. Like the ‘Launch Manager’ tab, the ‘Export Manager’ tab uses an outline view. All completed runs are listed.

Clicking the ‘+’ sign next to a Batch ID expands the item to reveal a list of the completed runs within it. Clicking the ‘+’ sign next to a Run ID expands the item to show the run’s characteristics. Figure 7-1 shows an example of the ‘Export Manager’ tab where the “EDDIES\_STATION\_B” batch and run 99 are expanded.

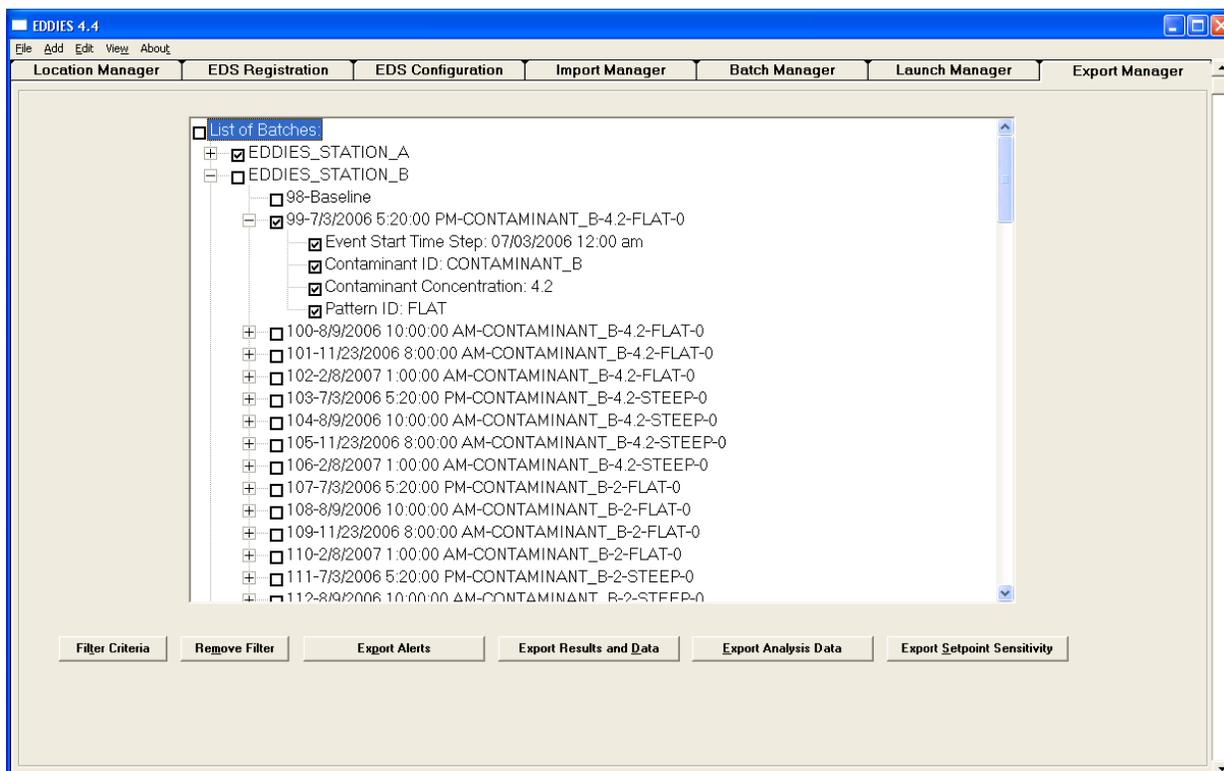


Figure 7-1. ‘Export Manager’ Tab

This section describes the export types. The format of these files are given in Appendix C.

Generating export files can take many hours. After initiating the export, expect to leave the workstation and return at a later time to collect the exported files, particularly if multiple runs or batches are selected.

### 7.1 Selecting Runs to Export

The procedure for selecting the run(s) to include in each export on the ‘Export Manager’ tab is the same regardless of export type. To select a run to include in the export, click the checkbox next to the Run ID so that a checkmark appears in it. Clicking the box next to a Batch ID selects all runs in the batch. Clicking the box next to the “List of Batch IDs” item selects all batches and, thus, all runs within the batches. Runs and batches can be unselected by clicking the box again so that the checkmark is removed. In Figure 7-1, the “EDDIES\_STATION\_A” batch has been selected (and thus all runs within that batch,

though not shown in an expanded view, will be included) and run 99 in the “EDDIES\_STATION\_B” batch have been selected.

### 7.1.1 Filtering

Selecting runs for export that meet specific criteria can be achieved efficiently through the filter capability on the ‘Export Manager’ tab. This is especially useful for the ‘Analysis Data’ export because it is often desired to analyze a subset of completed runs, such as all runs using a particular EDS configuration or all runs simulated using a given contaminant.

To select runs for an export through the filtering capability, select the ‘Filter Criteria’ button on the ‘Export Manager’ tab. The window containing all the run variables, shown in Figure 7-2, appears.

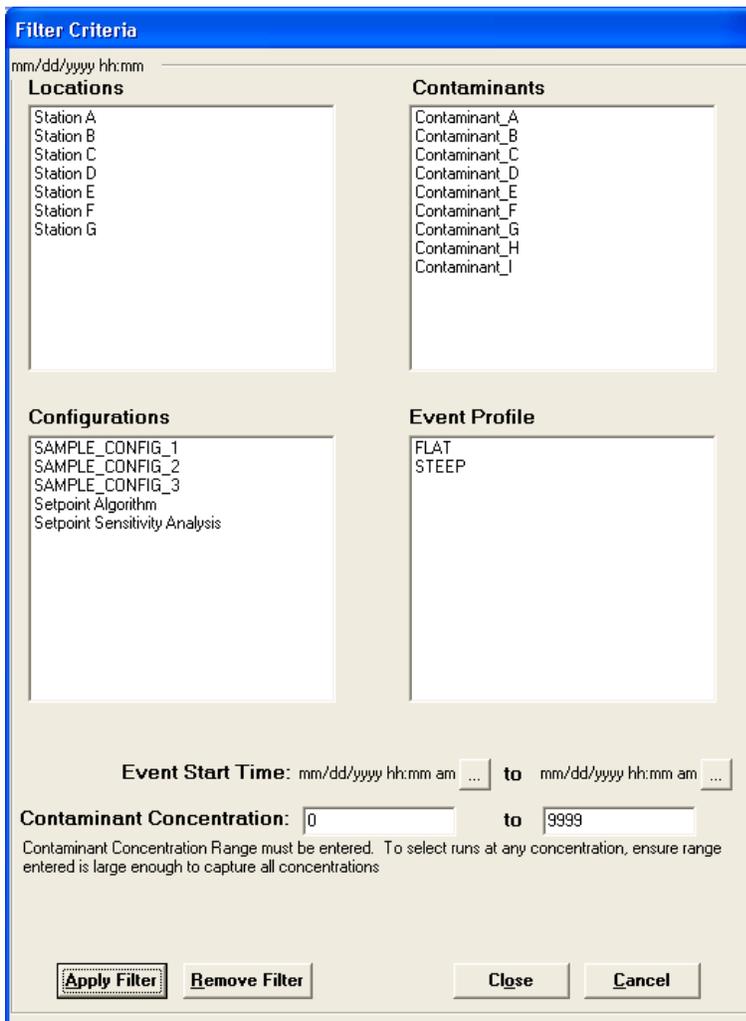


Figure 7-2. ‘Export Manager’ Tab Filter

The ‘Locations’, ‘Contaminants’, ‘Configurations’ and ‘Event Profile’ listboxes present all items defined within EDDIES, even if they are not used in a run. Any number of items within each listbox can be selected. Select an item by clicking on it so it is highlighted in blue. Unselect an item by clicking on it a second time so that it is no longer highlighted in blue.

The 'Event Start Time' and 'Contaminant Concentration' fields are not pre-defined. The filter is applied to these variables by entering the desired minimum and maximum values. The ranges are inclusive; thus the runs containing the values entered are included.

The 'Apply Filter' button removes all runs that do not match the entered criteria from the 'Export Manager' tab's outline view. The 'Remove Filter' button clears all filtering so that all completed runs are again shown in the 'Export Manager' tab's outline view.

## 7.2 Export Types

The 'Export Manager' tab generates four different types of export files: 'Alerts', 'Results and Data', 'Analysis Data' and 'Setpoint Sensitivity'. The contents and export procedure for each file type are detailed in this section.

### 7.2.1 Alerts Export

'Alerts' files list the alerts occurring in the selected run(s) according. Logic for determining alerts is described in Appendix D. Alerting can be determined using the ALERT\_STATUS or the ALERT\_LEVEL. These terms are defined in Section 1.2.

#### 7.2.1.1 File Contents

Table 7-1 shows an example alerts export file. The contents are further described in Section C.3.

**Table 7-1. Sample 'Alerts' Export**

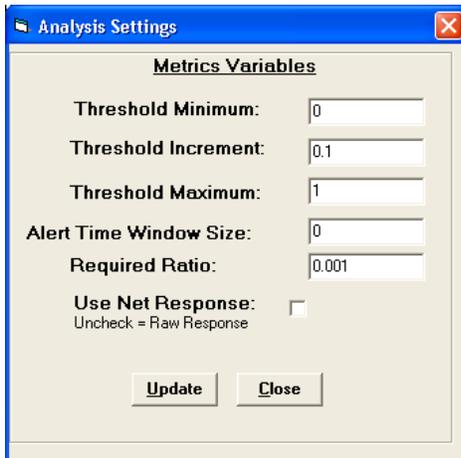
Analysis by Alert Status				
Alert Time Window=0				
RUN_ID	ALERT_START	ALERT_END	ALERT_COMMENT	
1	10/14/2007 13:00	10/15/2007 11:50	False Positive	
1	10/15/2007 15:20	10/15/2007 15:35	False Positive	
1	10/23/2007 12:15	10/23/2007 12:45	True Positive	Baseline Event Detected
1	12/20/2007 10:50	12/20/2007 10:55	False Positive	
2	11/5/2007 10:10	11/5/2007 12:10	True Positive	Simulated Event Detected
3	12/25/2007 14:50	12/25/2007 16:35	True Positive	Simulated Event Detected

#### 7.2.1.2 Procedure for Exporting

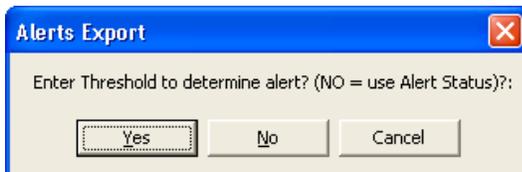
The following steps should be followed to export an 'Alerts' file.

1. From the 'Edit' menu, select the 'Analysis Variables' button and verify the alert time window in the screen that pops up (shown in Figure 7-3). This variable is described in Appendix D. The other variables in this window do not impact the 'Alerts' export calculations.
2. Select the run(s) to be exported from the 'Export Manager' tab as described in Section 7.1.
3. Select the 'Export Alerts' button at the bottom of the 'Export Manager' tab.
4. On the 'Enter/Select CSV File' browser that pops up, navigate to the desired directory of export, enter a file name, and click 'Open'. It is helpful to give the file a descriptive name.
5. On the 'Alerts Export' window that pops up (shown in Figure 7-4), define how alerts should be determined. Alert determination is discussed further in Appendix D.
  - Select 'No' to identify alerts using the ALERT\_STATUS (timesteps are alerting if ALERT\_STATUS = 1).

- Select 'Yes' to identify alerts using the ALERT\_LEVEL (timesteps are alerting if ALERT\_LEVEL  $\geq$  the threshold entered in Step 6).
6. If 'Yes' is selected, an 'Enter Threshold Value' window will pop up. Enter the desired discrimination threshold and click 'OK.'
  7. On the 'Export Results/Samples' window that pops up, specify the timestep format to be applied in the 'ALERT\_START' and the 'ALERT\_END' metrics of the export file.
    - Select 'Yes' for 12-hr timestep format (MM/DD/YYYY hh:mm am).
    - Select 'No' for 24-hour format (MM/DD/YYYY hh:mm).



**Figure 7-3. Analysis Settings**



**Figure 7-4. Alert Identification Method**

A message window will pop up when the export is complete.

## **7.2.2 Results and Data Export**

The 'Results and Data' export produces a CSV file in time-series format for each selected run. Each file consists of the EDS run results and, if the user chooses, the corresponding EVENT\_STATUS and the data presented to the EDS. Note that this export type always includes the EDS output. Time-series files that contain data only can be exported via the 'Launch Manager' tab as described in Section 6.2.2.

'Results and Data' export files are automatically named by Run ID. If the results are exported with data, the file is named 'Run\_X\_DataWithResults.csv' where X is the Run ID. If only results are exported, then the file is named 'Run\_X\_Results.csv'.

### **7.2.2.1 File Contents**

A sample 'Results and Data' export file containing the EVENT\_STATUS and data presented to the EDS is shown in Table 7-2. The contents are further described in Section C.4.

**Table 7-2. Sample ‘Results and Data’ Export**

Time_Step	D_ CL2	D_ PH	D_ TOC	ALERT_ STATUS	ALERT_ LEVEL	ANALYSIS_ COMMENTS	CONTRIBUTING_ PARAMETERS	EVENT_ STATUS
8/2/2008 0:00	1.07	8.89	1.35	0	0			0
8/2/2008 0:02	1.07	8.88	1.35	0	0			0
8/2/2008 0:04	1.07	8.88	1.35	0	0			0
8/2/2008 0:06	1.07	8.88	1.35	0	0			0
8/2/2008 0:08	1.07	8.88	1.35	0	0			0
8/2/2008 0:10	1.07	8.87	1.35	0	0			0
8/2/2008 0:12	1.06	8.89	1.35	0	0			0
8/2/2008 0:14	0.97	8.91	1.35	1	1		CL2,PH	0
8/2/2008 0:16	0.96	8.92	1.35	1	1		CL2,PH	0
8/2/2008 0:18	0.88	8.94	1.35	1	1		CL2,PH	0
8/2/2008 0:20	0.82	8.96	1.35	1	1		CL2,PH	0
8/2/2008 0:22	0.75	8.97	1.35	1	1		PH	0

### 7.2.2.2 Procedure for Exporting

‘Results and Data’ files can be created in two ways: during batch execution or using the ‘Export Manager’ tab. Generation of these files using the second method can take a long time.

#### **Exporting Files during Batch Execution**

This option must be selected before the batch is launched by checking the ‘Log Results and Data’ checkbox at the bottom of the ‘Launch Manager tab’ (as discussed in Section 6.2). Doing so opens the ‘Select Output Folder’ browser so a directory can be chosen in which to save CSV files.

One file is created for each run in the batch, including the baseline run, as the batch is executed. These export files include the data, EVENT\_STATUS, and EDS output and use the 12-hour timestep format. Each file is named ‘Run\_X\_DataWithResults.csv’ where X is the numeric Run ID.

#### **Export Using the ‘Export Manager’ Tab**

After batch execution, ‘Results and Data’ files can be exported for any or all of the completed runs using the ‘Export Manager’ tab as follows:

1. Select the run(s) to be exported as described in Section 7.1.
2. Click the ‘Export Results and Data’ button.
3. In the browser window that pops up, select a directory to write the export files and click ‘OK.’
4. On the window that pops up, specify if the data should be included in the file(s). Note that when using this method, the export files are created more quickly if data is not included.
  - Click the ‘Yes’ button to include data. If data is included, files are named ‘Run\_X\_DataWithResults.csv’ where X is the numeric Run ID.
  - Click the ‘No’ button to exclude data. If data is excluded, files are named ‘Run\_X\_Results.csv’ where X is the numeric Run ID.
5. On the next window, specify if the EVENT\_STATUS should be included in the file(s).
6. On the final ‘Export Results/Samples’ window that pops up, specify the timestep format to be applied in TIME\_STEP field of the export.
7. The progress of the export for each run is shown on the bottom of the ‘Export Manager’ tab in a status bar that appears when the export begins. A message pops up when the export is complete.

### **7.2.3 Analysis Data Export**

The ‘Analysis Data’ export presents performance metrics calculated from the EDS results at several alert thresholds based on the ALERT\_LEVEL. These exports are useful when choosing an alert threshold for deployment of an EDS. Metrics include the number of invalid alerts, the percentage of events detected, and average times to detect. Samples of each file type are shown in Appendix C, and calculation of the metrics is thoroughly explained in Appendix D. Note that the ALERT\_STATUS is not used in any way when calculating these metrics.

There are two options for ‘Analysis Data’ export files. The basic ‘Analysis Data’ export file contains standard metrics intended to meet the needs of the average user. The detailed ‘Analysis Data’ file contains additional, more complex metrics of interest to a more advanced user.

#### **7.2.3.1 Procedure for Exporting**

The steps below should be followed to produce an ‘Analysis Data’ file. Note that generation of these files can take a long time.

1. From the ‘Edit’ menu, select the ‘Analysis Variables’ button and verify the variable values (shown in Figure 7-3). These are described in Appendix D.
2. Select the run(s) to be exported from the ‘Export Manager’ tab as described in Section 7.1. For any batch from which an event run is selected, the baseline run must also be selected.
3. Select ‘Export Analysis Data’ button on this tab.
4. On the ‘Trigger Accuracy’ window that pops up, specify if the trigger accuracy metric should be included in the analysis file. This metric is only meaningful if the EDS results contain trigger parameters (described in Section 1.2).
5. On the ‘Report Detail Level’ window that pops up, select the desired report type.
  - Click ‘Yes’ for the basic ‘Analysis Data’ export file.
  - Click ‘No’ for the detailed ‘Analysis Data’ export.
6. On the browser window that pops up, navigate to the desired directory, enter a file name and click ‘Open’. It is helpful to give the file a descriptive name.

The progress of the export is shown across the bottom of the ‘Export Manager’ tab. Two windows will pop up successively when the export is complete. The first shows the export completion time and the second simply gives an indication of success.

### **7.2.4 Setpoint Sensitivity Export**

The ‘Setpoint Sensitivity’ export presents the performance of the setpoint sensitivity analysis on a single parameter. The performance of the setpoint sensitivity analysis is evaluated by applying the range of setpoint values as the minimum or maximum allowable parameter value. It is intended to assist users with selecting setpoint values to use in everyday operation.

As analysis using this functionality is quite different than that evaluating EDSs, a description of this file’s contents and the procedure for exporting this file type are given in Section E.2.

## Glossary

**Alert level.** A real number reflecting the EDS's assessment of the likelihood that conditions are anomalous, with higher values indicating more certainty that a water quality anomaly is occurring.

**Alert status.** A binary normal/abnormal indication for water quality outputted by an EDS which precisely identifies when the EDS is alerting.

**Application.** Synonymous with EDS.

**Baseline Dataset.** A set of raw utility data provided to the EDS. Each batch has one baseline dataset, and the data in this file depends on the location, date range and polling interval specified for the batch.

**Baseline Event.** A user-specified period of abnormal water quality in the utility data. It is expected that an EDS should alert during baseline events.

**Batch.** A set of *runs* that are executed together. The location and EDS configuration are the same for all runs in a batch.

**Alert.** A single notification of a potential water quality anomaly. An alert may include multiple alerting timesteps. See Appendix D for a precise definition.

**Alerting Timestep.** A timestep for which an EDS identifies a potential water quality anomaly. See Appendix D for a precise definition.

**Alert Time Window Size.** The minimum number of timesteps required between two alerting timesteps before they are classified as separate alarms.

**Contaminant.** A set of reaction expressions that relate the changes in a water quality parameter's value to the contaminant concentration.

**CSV File.** Stands for Comma Separated Value. A file in which each row's entries are separated by commas. Can be viewed easily in either Notepad or Excel.

**Dataset.** A set of data provided to an EDS at one time. A dataset is created for each run. Depending on the EDS type, this can be conveyed through Oracle database tables or a CSV file. The data contained in a dataset depends on the run type, location, data range, polling interval and simulated event characteristics (for event runs).

**Discrimination Threshold.** If the alert level outputted by the EDS is larger than this value, the tool is considered to be alerting.

**DMP File.** A file containing data from an Oracle database. In EDDIES, DMP files is created using the export.bat script and imported using the import.bat script.

**EDDIES (Event Detection, Deployment, Integration and Evaluation System).** An off-line software tool developed by the EPA to facilitate implementation of online water quality monitoring

**EDS (Event Detection System).** A software tool that analyzes water quality data and identifies for each timestep whether the data appears normal or abnormal.

**EDS Configuration.** A particular set of values for an EDS's configuration variables.

**EDS Configuration Variable.** A value that controls how an EDS executes during analysis. In general, each EDS has its own unique set of EDS configuration variables.

**EDS Response Time.** The length of time EDDIES will wait for acknowledgement from an EDS after posting a message before aborting the batch.

**Event.** A period of abnormal water quality. Baseline events occur in the raw utility data, and simulated events are generated by EDDIES. An EDS should alarm when presented with an event.

**Event Dataset.** A dataset containing one simulated event.

**Event Profile.** A time series of values, each containing a decimal indicating the percentage of the contaminant's peak concentration that is simulated for the given timestep. Defines the wave of contaminant that passes by the monitoring station.

**Event Timestep.** A timestep within an event – either a baseline event or simulated event.

**Event Start Timestep.** The time at which a simulated event begins.

**Location.** A group of parameters that is analyzed together by the EDS, usually associated with one monitoring station.

**Parameter.** A data stream. For example, one parameter might be the chlorine value from Station\_A, and another might be the sensor alarm associated with a sensor at Station\_K.

**Parameter Type.** What the parameter measures. This can be a water quality parameter type such as pH or chlorine, an alarm type or an operational data type such as a pump status or tank level.

**Polling Interval.** The frequency of data provided to the EDS to process. For example, if the polling interval were 5 minutes, data would be available at 12:00, 12:05, 12:10 and so on. This is also the interval for which results are collected.

**Reaction Factor.** An expression that relates the concentration of contaminant to the change in a given water quality parameter. See Appendix B for more details.

**Required Ratio.** The minimum ratio of alerting timesteps to the total number of timesteps in an event in order for the EDS to be credited with detecting the event.

**Run.** Defined by a testing dataset and EDS configuration.

**Run ID.** An automatically generated ID associated with a run.

**Simulated Event.** A contamination event simulated by EDDIES by superimposing water quality changes on uploaded data.

**Utility Data.** Data uploaded into EDDIES by the user.

# Appendix A: Import File Formats

As described throughout this document, data can be added to the EDDIES database by importing CSV files. Some important points about import file creation and management are:

- The first row of each import file *must* match the header rows listed in this appendix, in the order they are listed.
- The tables in this Appendix display the information in columns, though the required CSV format is text delineated by commas. As the comma-based format is difficult to work with as text, it is recommended that users populate and edit the information using Excel (select “CSV” as the file type when saving the file) or another CSV-editing software (many are available for free online).
- CSV file imports will fail if the file contains data for entries that are already in the database (i.e. the user tries to add a contaminant that is already in the contaminant library). The duplicate rows should be deleted or renamed to a new ID and then the file can be imported.

This appendix describes the required format for each import file type. The file types are presented in the order in which they would typically be imported into EDDIES. A sample file of each type can be found in the ‘Sample Import Files’ folder of the EDDIES directory.

## A.1 Parameter Type Import

Each row in the parameter type import file defines a parameter type. Parameter type import files have a fixed header row as follows: PARAMETER\_TYPE\_ID, PARAMETER\_TYPE\_DESCRIPTION, MINIMUM\_VALUE, MAXIMUM\_VALUE.

The remaining rows have the following values.

- PARAMETER\_TYPE\_ID: alphanumeric string with no spaces or commas
- PARAMETER\_TYPE\_DESCRIPTION: descriptive alphanumeric string that does not contain commas. Spaces are allowed.
- MINIMUM\_VALUE: any number; the minimum valid value for the parameter type. This field can be left blank. Section B.2.5 gives an example of how this is used.
- MAXIMUM\_VALUE: any number; the maximum valid value for the parameter type. This field can be left blank. Section B.2.5 gives an example of how this is used.

Table A-1 below shows the contents of sample file for parameter type import.

**Table A-1. Sample Parameter Type Import CSV File**

PARAMETER_TYPE_ID	PARAMETER_TYPE_DESCRIPTION	MINIMUM_VALUE	MAXIMUM_VALUE
CL2	Free Chlorine	0	
COND	Conductivity	200	600
PH	pH	5	
FLOW	Influent or Effluent Flow		

## A.2 Parameter Import

Each row in the parameter import defines a parameter. Parameter import files have a fixed header row as follows: PARAMETER\_ID, PARAMETER\_NAME, PARAMETER\_TYPE\_ID, SETPOINT\_LO, SETPOINT\_HI.

The remaining rows have the following values.

- PARAMETER\_ID: alphanumeric string with no spaces or commas
- PARAMETER\_NAME: descriptive alphanumeric string that does not contain commas. Spaces are allowed.
- PARAMETER\_TYPE\_ID: a parameter type already defined in EDDIES.
- SETPOINT\_LO: any number; the desired low setpoint value for the parameter (see Appendix E for more details). This field is can be left blank.
- SETPOINT\_HI: any number; the desired high setpoint value for the parameter (see Appendix E for more details). This field is can be left blank.

Table A-2 below shows the contents of a sample file for parameter import.

**Table A-2. Sample Parameter Import CSV File**

PARAMETER_ID	PARAMETER_NAME	PARAMETER_TYPE_ID	SETPOINT_LO	SETPOINT_HI
A_CL2x_V	Station A Chlorine	CL2	0.2	2
A_COND_V	Station A Conductivity	COND	300	450
A_PH_V	Station A pH	PH		
B_FLOW_OP	Station B Influent Flow	FLOW	0	

## A.3 Location Import

This file contains one row for each location to be imported to EDDIES. Location import files have a fixed header row as follows: LOCATION\_ID, LOCATION\_NAME.

The remaining rows have the following values.

- LOCATION\_ID: alphanumeric string with no spaces or commas.
- LOCATION\_NAME: descriptive alphanumeric string that does not contain commas. Spaces are allowed. This is the location name seen in EDDIES.

Table A-3 shows the contents of sample file for location import.

**Table A-3. Sample Location Import CSV File**

LOCATION_ID	LOCATION_NAME
TANK1	East Park Tank
RES_B	Reservoir B

## A.4 Location Parameter Import

Each row of the location parameter import file maps a parameter to a monitoring location. Parameters can be mapped to more than one location. Location parameter import files have a fixed header row as follows: LOCATION\_ID, PARAMETER\_ID

The remaining rows have the following values.

- LOCATION\_ID: a LOCATION\_ID previously defined in EDDIES.
- PARAMETER\_ID: a parameter previously defined in EDDIES.

Table A-4 shows the contents of a sample file for location parameter import.

**Table A-4. Sample Location Parameter Import CSV File**

LOCATION_ID	PARAMETER_ID
TANK1	T1_CL2x_V
TANK1	T1_COND_V
TANK1	FLOW_LINE1
RES_B	B_CL2x_V

## A.5 Utility Data Import

There are two acceptable formats for importing baseline data into EDDIES: time-series and parameter-based formats. The format of each is described below. Baseline data will not be overwritten; if an imported CSV file contains values for a parameter ID and timestep already in the database, the import will abort and an error message will be generated indicating that the baseline data already exists.

EDDIES supports both a 12-hour time format and 24-hour time format. An example of 12-hour format is 3/21/2009 8:00 pm. The same date/time in 24-hour format would be 3/21/2009 20:00.

The event status value for each timestep is defined by the user in the utility data to import. This field is critical to the results analysis, as described in Appendix D.

### A.5.1 Time-series Format

In the time-series format, each row can include values for multiple parameters for a single timestep. This format is primarily used for manually prepared data. Time-series files have the following header row:

- TIME\_STEP: this exact string.
- EVENT\_STATUS: this exact string.
- PARAMETER\_IDS: any PARAMETER\_ID previously defined in EDDIES. Each parameter should be written in a separate column. Any number of parameters can be included in one file, including parameters from multiple monitoring locations.

The remaining rows in a time-series import file should have the following values:

- TIME\_STEP: A string in one of the following formats: 12-hour (MM/DD/YYYY hh:mm am) or 24-hour time format (MM/DD/YYYY hh:mm).
- EVENT\_STATUS: -1, 0 or 1, as described in Appendix D.

- **PARAMETER\_IDS:** parameter values. Though generally numeric, these fields are designated as strings and thus anything can be entered. The parameter values must be in the order the IDs were listed in the header row.

Table A-5 shows the contents of sample file for baseline data import, in time-series format. This example uses the 24-hour time format.

**Table A-5. Sample Baseline Data Import CSV File in Time-series Format**

TIME_STEP	EVENT_STATUS	A_CL2_V	A_COND_V	A_PH_V
1/2/2008 13:22	0	1.07	282.75	8.51
1/2/2008 13:24	0	1.11	282.32	8.51
1/2/2008 13:26	0	1.09	282.44	8.52

### A.5.2 Parameter-based Format

In this parameter-based format, there can be multiple rows for each timestep, with each row containing data for a single parameter for a timestep. Data from a historian or data management system is often in this format. For parameter-based files, there is a fixed header row as follows: **PARAMETER\_ID**, **EVENT\_STATUS**, **TIME\_STEP**, **SAMPLE\_VALUE**, **SAMPLE\_QUALITY**.

The remaining rows have the following values:

- **PARAMETER\_ID:** the **PARAMETER\_ID** of the data to be reported. This must have been previously defined in EDDIES.
- **EVENT\_STATUS:** -1, 0 or 1, as described in Appendix D.
- **TIME\_STEP:** a string in one of the following formats: 12-hour (MM/DD/YYYY hh:mm am) or 24-hour time format (MM/DD/YYYY hh:mm).
- **SAMPLE\_VALUE:** parameter value. Though generally numeric, this field is designated as a string and thus anything can be entered.
- **SAMPLE\_QUALITY:** 'Normal' or 'Bad'. This field can be left blank, though the column and column heading must be included in the CSV. **SAMPLE\_QUALITY** is not used by EDDIES; it is simply passed on to the EDS, which may or may not use the information.

Table A-6 shows the contents of sample file for baseline data import, in parameter-based format. This example uses the 12-hour time format.

**Table A-6. Sample Baseline Data Import CSV File in Parameter-based Format**

PARAMETER_ID	EVENT_STATUS	TIME_STEP	SAMPLE_VALUE	SAMPLE_QUALITY
A_CL2_V	0	1/2/2008 1:22 PM	1.07	
A_COND_V	0	1/2/2008 1:22 PM	282.75	
A_PH_V	0	1/2/2008 1:22 PM	8.51	
A_CL2_V	0	1/2/2008 1:24 PM	1.11	Normal
A_COND_V	0	1/2/2008 1:24 PM	282.32	Normal
A_PH_V	0	1/2/2008 1:24 PM	8.51	Normal
A_CL2_V	0	1/2/2008 1:26 PM	0	Bad
A_COND_V	0	1/2/2008 1:26 PM	282.44	Normal
A_PH_V	0	1/2/2008 1:26 PM	8.52	Normal

In parameter-based format, an event status is required for each parameter, and thus multiple event status values are entered by the user for each timestep. EDDIES combines the event statuses for all parameters assigned to a location to get the overall event status for that timestep. Event status values are given a

priority of 1, -1 and 0. Thus, if any parameter imported using a parameter-based format has an event status of 1, EDDIES assigns the entire timestep an event status of 1. If no parameters have an event status of 1 but at least one has an event status of -1, the timesteps is assigned an event status of -1. If all parameters have an event status of 0, the overall event status is 0.

## A.6 EDS Registration Import

For EDS registration using CSV import, two types of files need to be imported. First, an ‘EDS’ file and then an ‘EDS Properties’ file. The descriptions of both file types are provided in the sections below.

### A.6.1 EDS Import

This file contains one row for each EDS and defines the EDS in EDDIES. EDS is synonymous with the term Application used in EDDIES. EDS import files have a fixed header row as follows: APPLICATION\_ID, APPLICATION\_NAME.

The remaining rows have the following values:

- APPLICATION\_ID: alphanumeric string with no spaces or commas.
- APPLICATION\_NAME: descriptive alphanumeric string that does not contain commas. This field can have spaces and lower-case letters.

Table A-7 shows the contents of a sample file for EDS import.

**Table A-7. Sample EDS Import CSV File**

APPLICATION_ID	APPLICATION_NAME
CANARY	CANARY
TOOL_B	EDS Tool B

### A.6.2 EDS Property Import

EDS property import files have a fixed header row as follows: APPLICATION\_ID, PROPERTY\_NAME, PROPERTY\_VALUE.

The remaining rows have the following values. The user should reference the EDS documentation for guidance on populating these fields for a specific EDS.

- APPLICATION\_ID: an APPLICATION\_ID previously defined in EDDIES.
- PROPERTY\_NAME: ‘AppDirectory’, ‘AppFileName’, ‘DataDirectory’, ‘Contact’, ‘Email’ or ‘Phone’. AppDirectory, AppFileName and DataDirectory are required fields. Contact information can be left blank.
- PROPERTY\_VALUE: any alphanumeric string that does not contain commas. Can be left blank for the contact information fields

Table A-8 shows the contents of a sample file for EDS property import.

**Table A-8. Sample EDS Property Import CSV File**

APPLICATION_ID	PROPERTY_NAME	PROPERTY_VALUE
EDS_A	AppDirectory	C:\ProgramFiles\EDS_A
EDS_A	AppFileName	eds_a.exe
EDS_A	DataDirectory	C:\EDDIESData\EDS_A
EDS_A	Contact	John Brown
EDS_A	Email	
EDS_A	Phone	

## A.7 EDS Variable Import

The EDS Variable import file defines the configuration variables associated with a particular EDS(s). EDS variable import files have a fixed header row as follows: APPLICATION\_ID, VARIABLE\_NAME, VARIABLE\_DESCRIPTION, VARIABLE\_ACCEPTED\_VALUE, VARIABLE\_DEFAULT\_VALUE.

The remaining rows have the following values.

- APPLICATION\_ID: an APPLICATION\_ID previously defined in EDDIES.
- VARIABLE\_NAME: an alphanumeric string with no spaces.
- VARIABLE\_DESCRIPTION: a descriptive alphanumeric string that does not contain commas. Spaces are allowed. This field can be left blank.
- VARIABLE\_ACCEPTED\_VALUE: any alphanumeric string that does not contain commas. Spaces are allowed. This field can be left blank.
- VARIABLE\_DEFAULT\_VALUE: any alphanumeric string that does not contain commas. Spaces are allowed. This field can be left blank.

Table A-9 shows the contents of a sample file for EDS variable import.

**Table A-9. Sample EDS Variable Import CSV File**

APPLICATION_ID	VARIABLE_NAME	VARIABLE_DESCRIPTION	VARIABLE_ACCEPTED_VALUE	VARIABLE_DEFAULT_VALUE
EDS_A	WindowSize	Number of timesteps to consider in analysis.	Any whole number	
EDS_A	Threshold		Value between 0 and 1	1
EDS_A	Use_clusters	Yes or No		

## A.8 EDS Configuration Import

Each row of an EDS configuration import file contains the value for a single EDS variable. EDS configuration import files have a fixed header row as follows: CONFIGURATION\_ID, APPLICATION\_ID, VARIABLE\_NAME, VARIABLE\_VALUE.

The remaining rows have the following values. The user should reference the EDS documentation for required field formats.

- CONFIGURATION\_ID: an alphanumeric string with no spaces or commas.
- APPLICATION\_ID: an EDS\_ID previously defined in EDDIES.

- VARIABLE\_NAME: a configuration variable previously defined in EDDIES for the selected EDS.
- VARIABLE\_VALUE: any alphanumeric string that does not contain commas.

Table A-10 shows the contents of a sample file for EDS configuration import.

**Table A-10. Sample EDS Configuration Import CSV File**

CONFIGURATION_ID	APPLICATION_ID	VARIABLE_NAME	VARIABLE_VALUE
StationA_EDSA	EDS_A	WindowSize	100
StationA_EDSA	EDS_A	Threshold	0.9
StationA_EDSA	EDS_A	Use_clusters	No
StationB_EDSA	EDS_A	WindowSize	20
StationB_EDSA	EDS_A	Threshold	1.0
StationB_EDSA	EDS_A	Use_clusters	Yes

## A.9 Contamination Event Libraries Import

In order to define event runs in batches, the contaminant and event profile have to be predefined in EDDIES prior to creating batches. The sections below provide information on these import file formats.

### A.9.1 Contaminant Import

Each row of a contaminant import file contains the reaction expression for single contaminant and parameter impacted by the contaminant. Contaminant import files have a fixed header row as follows: CONTAMINANT\_ID, PARAMETER\_TYPE\_ID, REACTION\_FACTOR.

The remaining rows have the following values:

- CONTAMINANT\_ID: an alphanumeric string with no spaces or commas.
- PARAMETER\_TYPE\_ID: a parameter type previously defined in EDDIES.
- REACTION\_FACTOR: a string with no spaces containing the reaction factor. This field can include any number, letter or special character. See Section B.3.1 for guidance on developing reaction factors.
  - 'X' indicates the concentration variable. If an 'X' is not included, the reaction factor will be a constant value and the parameter will have that value for all event timesteps.
  - '+', '-', '\*' and '/' are valid operators.
  - Powers must be written using multiplication. For example,  $X^2$  must be written as 'X\*X'.

Table A-11 shows the contents of a sample file for contaminant import.

**Table A-11. Sample Contaminant Import CSV File**

CONTAMINANT_ID	PARAMETER_TYPE_ID	REACTION_FACTOR
ContaminantA	CL2	-0.3*X
ContaminantA	ORP	-8.5*X+25
ContaminantB	TOC	2.5*X*X

### A.9.2 Event Profile Import

Each row of an event profile import file contains the percent of peak concentration (see Section B.2 for details on percent of peak concentration) for a profile and event timestep. Event profile import files have a fixed header row as follows: EVENT\_PROFILE, PERCENT\_PEAK\_CONCENTRATION, STEP.

The remaining rows have the following values:

- EVENT\_PROFILE: an alphanumeric string with no spaces or commas.
- PERCENT\_PEAK\_CONCENTRATION: any positive real number, typically between 0 and 1.
- STEP: a whole number. The file must start with a step of 1 and increment by 1.

Table A-12 shows the contents of a sample file for event profile import.

**Table A-12. Sample Event Profile Import CSV File**

EVENT_PROFILE	PERCENT_PEAK_CONCENTRATION	STEP
PROFILE_A	0.1	1
PROFILE_A	0.2	2
PROFILE_A	1	3
PROFILE_A	0.5	4
PROFILE_A	0	5
PROFILE_B	2	1
PROFILE_B	4	2
PROFILE_B	1	3

### A.10 Batch and Run Import

Each row of a run import file contains the characteristics for one simulated contamination run. Defined batch(es) contain the exact runs specified in the file: EDDIES does not do any combination of entered run characteristics like it does when creating batches and runs in the GUI (as described in Section 6.1.2).

Run information is specified in the file. Overall batch information including the batch ID, location ID, start date, end date, configuration ID and polling rate are entered on pop-up screens when the run import CSV file is imported (see Section 6.1.2). Any contaminant or event profile used must already be in the EDDIES database. Run import files have a fixed header row as follows: BATCH\_ID, EVENT\_START\_TIME\_STEP, CONTAMINANT\_ID, CONTAMINANT\_CONCENTRATION, EVENT\_PROFILE.

The remaining rows have the following values:

- BATCH\_ID: an alphanumeric string with no spaces or commas.
- EVENT\_START\_TIME\_STEP: a string in one of the following formats: 12-hour (MM/DD/YYYY hh:mm am) or 24-hour time format (MM/DD/YYYY hh:mm).
- CONTAMINANT\_ID: a CONTAMINANT\_ID previously defined in EDDIES.
- CONTAMINANT\_CONCENTRATION: any real number.
- EVENT\_PROFILE: an EVENT\_PROFILE\_ID previously defined in EDDIES.

Table A-13 shows the contents of a sample file for batch and run import.

**Table A-13. Sample Batch and Run Import CSV File**

BATCH_ID	EVENT_START_TIME_STEP	CONTAMINANT_ID	CONTAMINANT_CONCENTRATION	EVENT_PROFILE
BATCH_A	6/10/2008 1:00	ContaminantA	0.5	FLAT
BATCH_A	6/10/2008 1:00	ContaminantA	2.5	STEEP
MY_BATCH_1	6/10/2008 1:00	ContaminantB	1.33	FLAT

## A.11 EDS Results Import

Section 6.2 describes how an EDS evaluation can be completed outside of EDDIES by outputting test datasets, running them through the EDS, and importing the results into the EDDIES database via a results file import. When importing results, there must be one CSV import file for each dataset. Each CSV results file must be named 'Run\_X\_Results.csv', where X is the run number. There should be a row in the results file for each timestep in the corresponding data file.

Results import files must have the following fixed header row: TIME\_STEP, ANALYSIS\_COMMENTS, ALERT\_STATUS, ALERT\_LEVEL, TRIGGER\_PARAMETERS.

The remaining rows have the following values. See Section 1.2 for a detailed description of the EDS output fields.

- TIME\_STEP: a string in one of the following formats: 12-hour (MM/DD/YYYY hh:mm am) or 24-hour time format (MM/DD/YYYY hh:mm). This timestep must have appeared in the corresponding dataset file.
- ANALYSIS\_COMMENTS: any alphanumeric string that does not contain commas. Spaces are allowed. Maximum field length is 1,000 characters.
- ALERT\_STATUS: typically a value of 0 or 1.
- ALERT\_LEVEL: any real number.
- TRIGGER\_PARAMETERS: an alphanumeric string. The entire contents of the field must be contained in double quotations.

Table A-14 shows the contents of a sample file for EDS results import.

**Table A-14. Sample EDS Results Import CSV File**

TIME_STEP	ANALYSIS_COMMENTS	ALERT_STATUS	ALERT_LEVEL	TRIGGER_PARAMETERS
1/1/2008 12:00 AM	Filling buffer	0	0.1	
1/1/2008 12:02 AM	Filling buffer	0	0.24	
1/1/2008 12:04 AM		0	0.1	
1/1/2008 12:06 AM		0	0.23	
1/1/2008 12:08 AM		0	0.16	
1/1/2008 12:10 AM		0	0.29	
1/1/2008 12:12 AM		0	0.39	
1/1/2008 12:14 AM	Anomaly detected	1	0.8	"CL"
1/1/2008 12:16 AM	Anomaly detected	1	0.9	"CL, COND"

## Appendix B: Dataset Generation

A *dataset* is a set of data presented to an EDS at one time for analysis. There are two types of datasets used in EDDIES: baseline datasets and event datasets. Both are generated based on the location, time period and polling interval specified by the user during batch creation. Baseline datasets use the raw data values imported by the user. Note that these datasets may contain user-defined baseline events, as discussed in Section D. Each event dataset contains one simulated contamination event, superimposed on the imported utility data based on user-specified event characteristics. As noted in Section 6.1, each batch contains one baseline run and one or more event runs, each using the same location, time period, polling interval and EDS configuration.

Section B.1 describes how the location, time period and polling interval are used to generate both types of datasets. Section B.2 describes in detail how simulated contamination events are generated, and Section B.3 describes how to view and modify the contaminant and event profile libraries used during event creation.

### B.1 Application of Location, Time Period and Polling Interval in Dataset Creation

For each batch, the user must specify a single location, time period and polling interval. EDDIES uses these values when generating all datasets for the batch.

It is possible that a user imported months of data for hundreds of parameters. Also, the data might or might not have a set polling interval (e.g., data for all parameters is provided for all even timesteps). All of this data is not simply “dumped on” the EDS during a batch. Instead, EDDIES pre-processes the data and posts (or exports) a dataset that contains only the data relevant for the batch for consistent timesteps.

#### B.1.1. Batch Location

As noted in Section 4, a location defines a group of parameters to be analyzed together. Commonly, data from all sensors co-located at a monitoring location are assigned to a location in EDDIES.

The monitoring location to be used is specified during batch creation. For all datasets in the batch, all parameters mapped to that location (and only those parameters) are included. Note that the parameters mapped to a location cannot be changed once a batch using that location has been implemented: a different set of parameters yields different datasets and thus the results in the database from the previous batch would be rendered invalid.

#### B.1.2. Batch Time Period

The user must define the time period to include in a batch’s datasets. This ensures that precisely the data the user wants to evaluate is included and that the same time period is included for all parameters.

Baseline datasets start at 12:00 am (0:00) on the start date specified during batch creation, and end on the last timestep of the end date according to the polling interval (see Section B.1.3 for more on this). All specified simulated events must fall within this same time period. The user is able to (and is encouraged to) use the Start EDS Analysis and End EDS Analysis variables described in Section 6.2.1.1 to provide the EDS only the amount of data needed for valid analysis (some EDS tools require a certain amount of data for training or buffering before they are able to provide valid output). This significantly reduces the time required to export the datasets and run the EDS.

### **B.1.3. Batch Polling Interval**

Many EDSs expect to receive values for all parameters at a set interval: the EDS receives one timestep of data, analyzes it, outputs results indicating if it finds the data normal or abnormal, and then repeats the process for the next timestep of data. This ensures the data is processed accurately and results are provided regularly, allowing for meaningful evaluation.

The polling interval in EDDIES ensures that all datasets have a regular interval appropriate for EDS analysis, even if the uploaded data does not have this consistency. Reasons that this is needed include that utility data may have different data collection or storage intervals across parameters, or there may be missing data due to communication issues or data export errors.

All datasets begin at 12:00 am (0:00). Baseline datasets begin on the batch start date, and event dataset start dates are based on the Start EDS Analysis value. Beginning at 0:00, EDDIES generates data at a frequency equal to the specified polling interval. For each timestep, the most recent value is reported for each parameter. If necessary and available, EDDIES uses utility data from before the start date to populate the first parameter values.

Table B-1 provides an example of utility data for one parameter, A\_PH. Tables B-2 and B-3 show examples of how EDDIES generates the baseline dataset for two different polling intervals – both using a batch Start Date of 2/1/2009. In each, the data from 1/31/2009 was used as no data was reported at 2/1/2009 12:00 am. Note that depending on the interval chosen, some data points from the imported utility data may be repeated, whereas others may not show up in the baseline dataset at all.

**Table B-1. Example A\_PH Baseline Data**

<b>Timestep</b>	<b>A_PH</b>
1/31/2009 11:57 PM	7.17
2/1/2009 12:06 AM	7.18
2/1/2009 12:07 AM	7.19
2/1/2009 12:15 AM	7.20

**Table B-2. Example A\_PH Baseline Data Transformed to a 2-Minute Polling Interval**

<b>Timestep</b>	<b>A_PH</b>
2/1/2009 12:00 AM	7.17
2/1/2009 12:02 AM	7.17
2/1/2009 12:04 AM	7.17
2/1/2009 12:06 AM	7.18
2/1/2009 12:08 AM	7.19
2/1/2009 12:10 AM	7.19
2/1/2009 12:12 AM	7.19
2/1/2009 12:14 AM	7.19
2/1/2009 12:16 AM	7.20
2/1/2009 12:18 AM	7.20
2/1/2009 12:20 AM	7.20

**Table B-3. Example A\_PH Baseline Data Transformed to a 10-Minute Polling Interval**

<b>Timestep</b>	<b>A_PH</b>
2/1/2009 12:00 AM	7.17
2/1/2009 12:10 AM	7.19
2/1/2009 12:20 AM	7.20

With the two-minute polling interval shown in Table B-2, there are many repeated values because the polling interval chosen is less than the frequency with which data is reported. This can lead to inaccurate

EDS analysis and an alert is often produced when the value changes. Thus, the polling interval chosen should be greater than or equal to the frequency at which data is reported.

This same procedure is repeated for all parameters assigned to the monitoring location – regardless of if they are reported at the same times. Table B-4 gives sample utility data from the same time period for another parameter at the same location.

**Table B-4. Example A\_CL2 Baseline Data**

Timestep	A_CL2
2/1/2009 12:00 AM	1.2
2/1/2009 12:05 AM	1.18
2/1/2009 12:10 AM	1.16
2/1/2009 12:15 AM	1.18

Assuming that A\_PH and A\_CL2 are the only parameters mapped to the given monitoring location, Table B-5 shows the data that would be in the baseline dataset if a 10-minute polling interval were selected by the user. The time-series format is used in this table.

**Table B-5. Resulting Data in the Baseline Dataset Using a 10-Minute Polling Interval**

Timestep	A_CL2	A_PH
2/1/2009 12:00 AM	1.2	7.17
2/1/2009 12:10 AM	1.16	7.19
2/1/2009 12:20 AM	1.18	7.20

## B.2 Simulation of Contamination Events

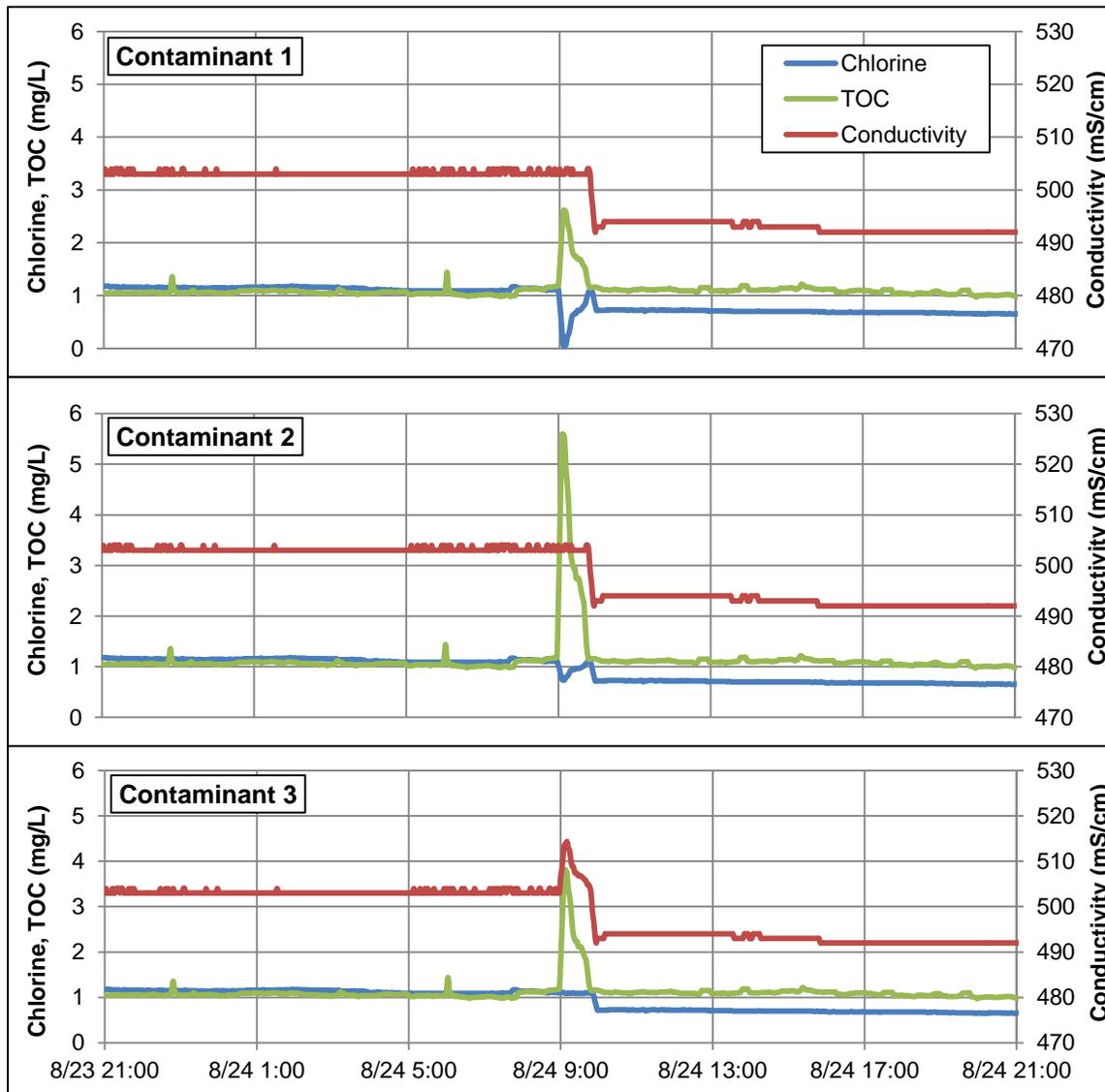
In EDDIES, contamination events are created by modifying utility water quality data in a manner that simulates contamination. Simulated contamination events are generated based on the location, polling interval and run properties defined by the user during batch creation, as discussed in Section 6.1. The run properties consist of the event start timestep, contaminant, peak contaminant concentration and event profile. In EDDIES, the event start time establishes the date and time at which the simulated contamination event begins, the contaminant specifies how and which baseline water quality values are modified, the concentration specifies the magnitude of the modification, and the event profile specifies the pattern in which the contaminant is seen over the length of the event.

This section explains how the run properties are used to calculate the water quality changes superimposed on the baseline dataset to generate simulated contamination events. Section B.2.5 demonstrates generation of a sample event.

### B.2.1. Contaminant

In EDDIES, contaminants are defined using reaction factors, which give the change in a water quality parameter in terms of contaminant concentration. The contaminant used determines which parameters are impacted and the relative impact across parameters.

Figure B-1 illustrates how the contaminant selected impacts the simulated event. Chlorine, conductivity and TOC are shown for simulated events using three different contaminants. The remaining run properties – location and time period, event start time, peak contaminant concentration and event profile – are held constant. The event period is 8/24 from 9:00 to 9:46.



**Figure B-1. Example Simulated Event Varying by Contaminant**

Contaminant 1 causes a decrease in chlorine and increase in TOC at about the same intensity. Contaminant 2 causes a very large change in TOC and a small change in chlorine. Contaminant 3 impacts TOC and conductivity. Note that the data outside of the event period is identical in all three events and is the same data as in the baseline dataset.

Table B-6 shows the default contaminants included with EDDIES, given as reaction expressions where “X” is the contaminant concentration. These reaction factors were derived from pipe-loop tests performed at EPA’s Testing and Evaluation Center (Hall, et. al, 2007). A variety of contaminants were injected at different strengths to test how the various water quality parameters were impacted. The tested contaminants were identified by the USEPA as contaminants of concern because of their availability, ability to be dispersed in water, and ability to cause harm if injected into a water system.

These default contaminants are loaded into the EDDIES database when the eddies\_defaults.dmp file is imported as described in Section 3.2.2.2. They can also be found in the ‘Sample Import Files’ folder within the EDDIES installation package.

**Table B-6. Default Contaminants in Terms of Contaminant Concentration**

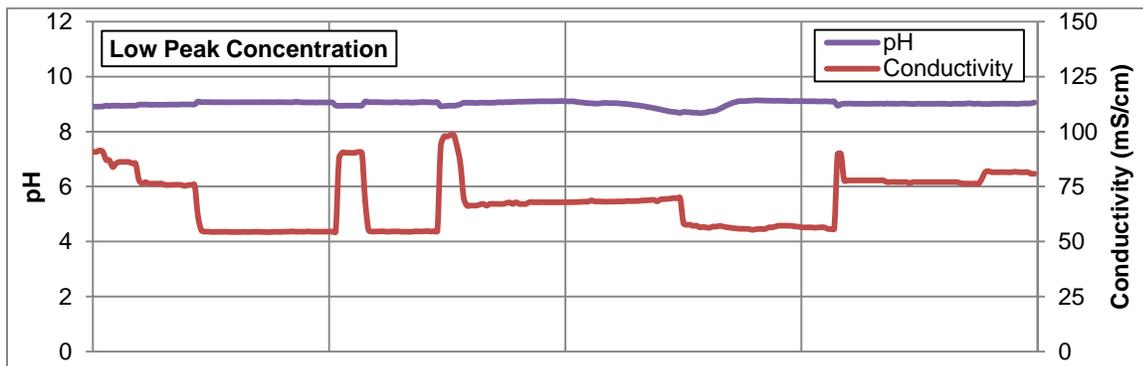
	CL2	COND	PH	TOC	ORP	UV
Contaminant_A	-0.02*X	-	0.02*X	0.05*X	-	-
Contaminant_B	-0.78*X	2.3*X	0.08*X	0.08*X	-1.8*X	-
Contaminant_C	-0.03*X	-	-	0.35*X	-	0.04*X
Contaminant_D	-0.24*X	-	-	-	-55*X	-
Contaminant_E	-0.07*X	-	0.01*X	0.57*X	-3.3*X	0.024*X
Contaminant_F	-0.002*X	0.19*X	-0.05*X	-	3.6*X	0.0024*X
Contaminant_G	-	0.78*X	-	0.24*X	-	-
Contaminant_H	-	1.99*X	-	-	13.8*X	-
Contaminant_I	-0.19*X	0.4*X	-	-	-50*X	-
Contaminant_J	-	-	-	0.66*X	-	-

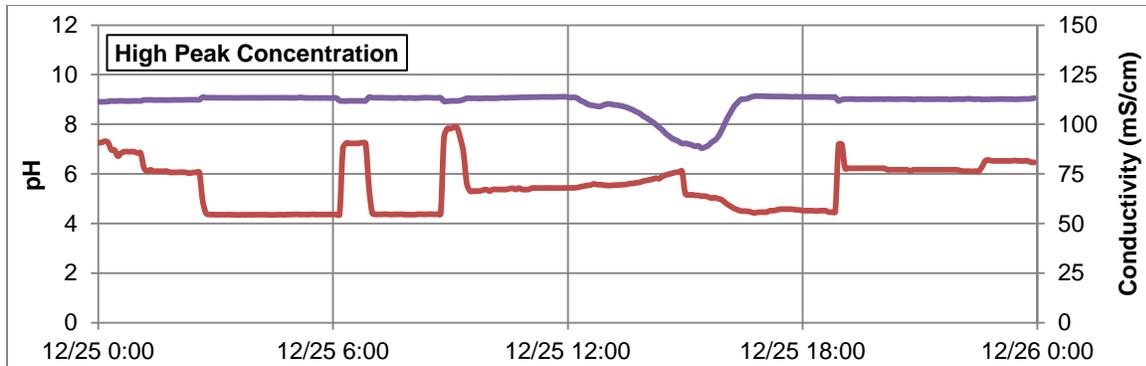
Contaminant(s) for a batch are selected during batch creation, described in Section 6.1. Contaminants must be defined in the EDDIES database before they can be selected for a batch. Section B.3 describes how to view and add contaminants.

**B.2.2. Peak Contaminant Concentration**

As described in the previous section, contaminants are defined in terms of contaminant concentration. The concentration at a given timestep is found by combining the event profile (described in Section B.2.3) with the peak concentration. The peak concentration is the maximum contaminant concentration that will occur during the event.

Figure B-2 show the differences in water quality changes in two simulated contamination events at two peak contaminant concentrations (low and high). All other run properties are the same for both events. The event period is 12/25 from 12:00 to 16:40.





**Figure B-2. Example Simulated Event with Low and High Peak Contaminant Concentrations**

The contaminant simulated in these events impacts pH strongly and has a weaker impact on conductivity. When the lower peak contaminant concentration is used, the water quality changes are hardly discernible (this event would likely not be detected by an EDS). The high peak contaminant concentration causes significant water quality changes, particularly in pH.

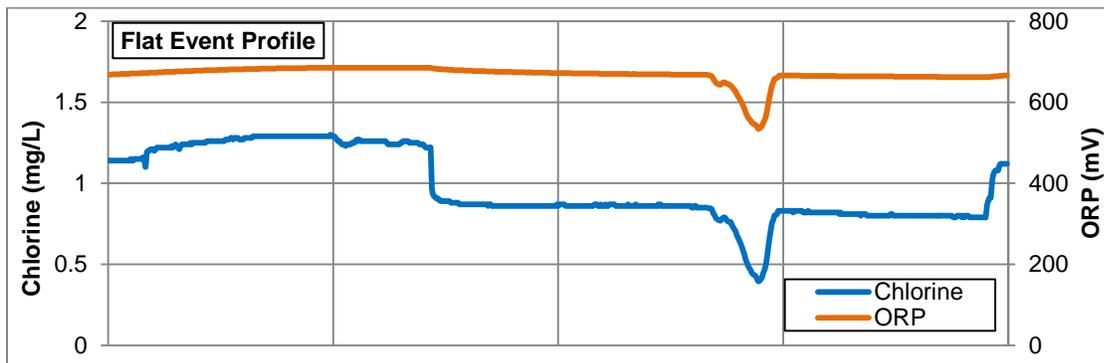
When selecting peak concentrations, it is suggested that the user plug values into the contaminant's reaction expressions until the maximum desired water quality change is achieved. For example, if a peak concentration of 10 were entered for Contaminant J, the largest change in TOC would be  $0.66 \times 10 = 6.6$ .

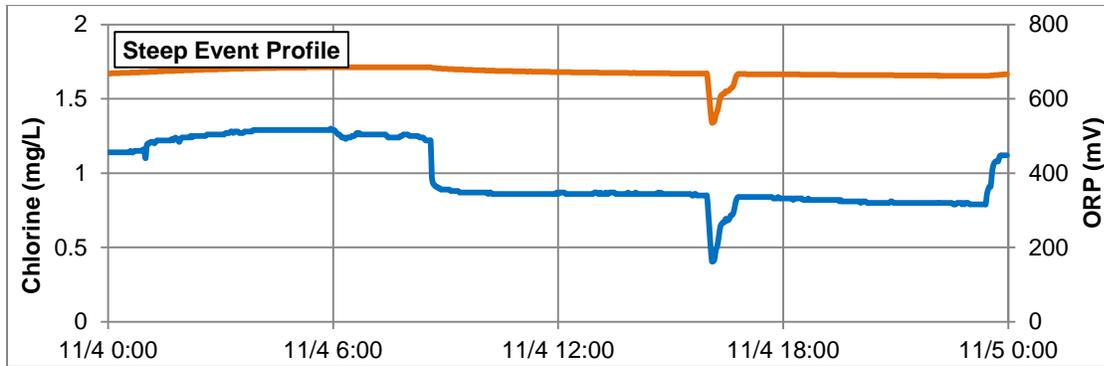
Peak contaminant concentration(s) for a batch are entered during batch creation, described in Section 6.1.

### **B.2.3. Event Profile**

As a contamination event would likely last for multiple timesteps with the concentration of contaminant at each timestep varying over that time period, the event profile is used to define the rise and fall of the contaminant concentration, as well as the length of the event. Event profiles are defined by a time series of percentages of the contaminant's peak concentration.

Figure B-3 shows the chlorine and ORP data resulting when two different event profiles were used when simulating contamination events. All other event characteristics are the same. Both events begin on 11/4 at 16:00, but the event length is different for the two profiles. The event with the flat event is longer and lasts until 17:52, whereas the steep event is over at 16:46.





**Figure B-3. Example Contamination Event with Flat and Steep Event Profiles**

Though the type and magnitude of water quality changes is the same in both events, they appear quite different. The flat profile causes a more gradual, though longer, change in water quality, whereas the steep profile causes an abrupt change in water quality.

The default event profiles included in EDDIES are shown graphically in Figure B-4. The default event profiles are also listed in the sample event profiles import file (EventProfiles.csv located in the 'Sample Import Files' folder within the EDDIES installation package).

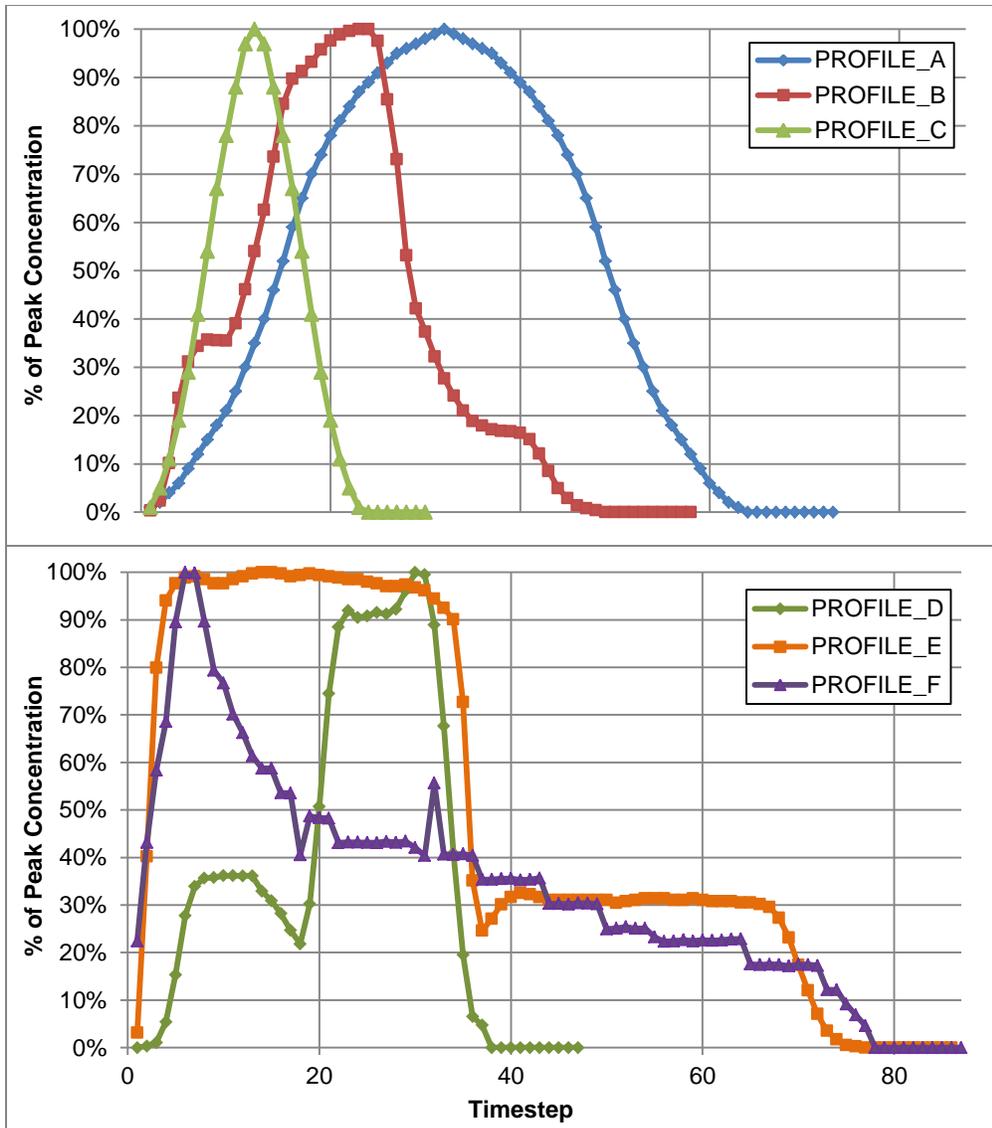


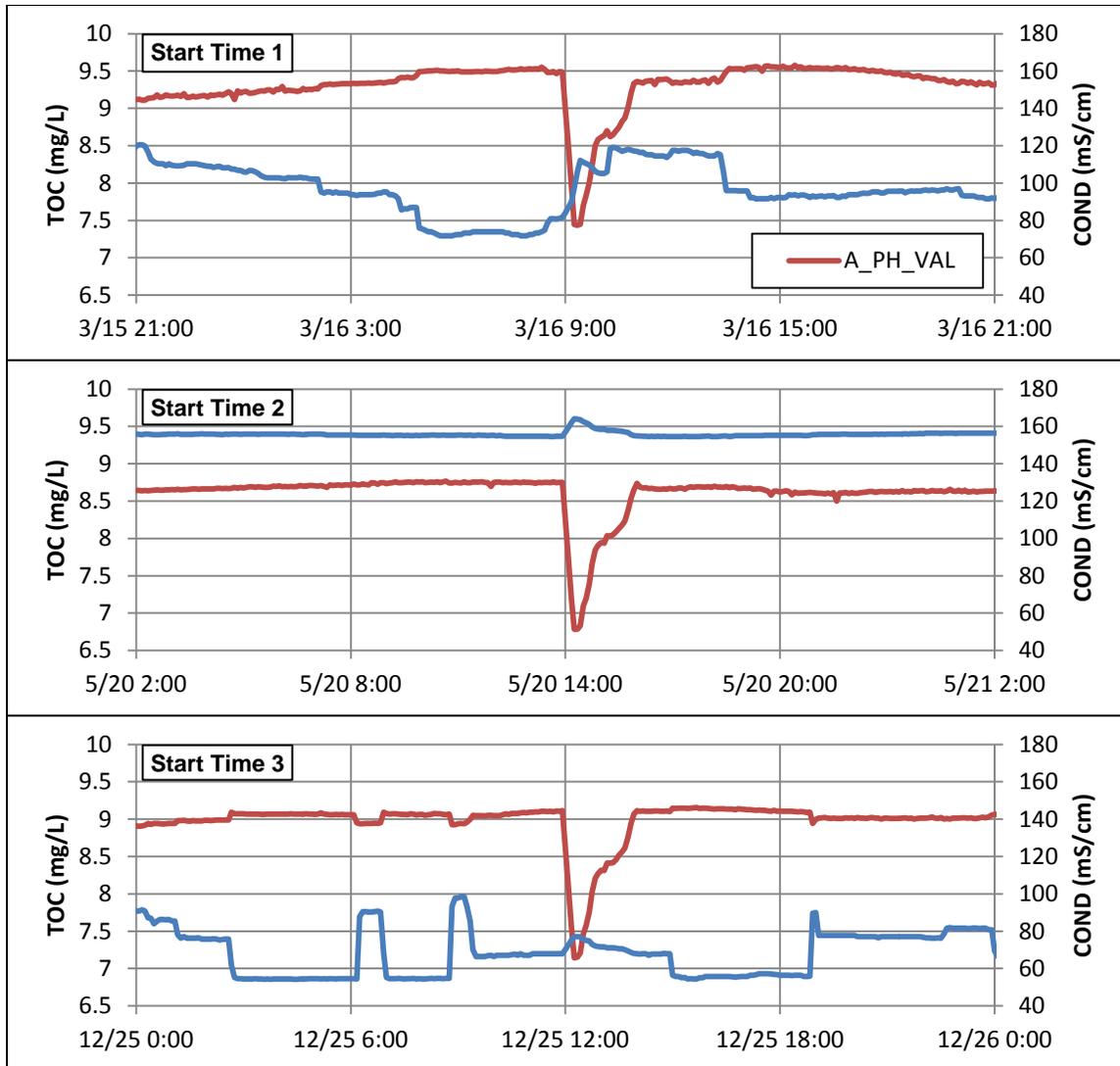
Figure B-4. Default Event Profiles A, B, C, D, E and F

Event profile(s) for a batch are selected during batch creation, described in Section 6.1. Event profiles must be defined in the EDDIES database before they can be selected for a batch. Section B.3 describes how to view and add profiles.

#### B.2.4 Event Start Time

The event start time establishes the date and time in the baseline data at which event simulation begins. The first timestep of the event profile is implemented at the event start time, and the remaining profile timesteps continue in sequence until all have been applied.

Figure B-5 shows examples of contamination events where the start time is varied. All other characteristics are held constant.



**Figure B-5. Example Simulated Contamination Event with Varying Start Times**

The drop in pH is obvious in all three events. However, the changes in conductivity appear very different due to water quality variability in the baseline data. With start time 1, the change in conductivity is entirely hidden by the change already occurring in the baseline data. It is clearly visible for start time 2, however, as the conductivity around that time is very stable. Also, note that the conductivity data values are quite different among the plots. It varies from about 7.3 to 8.5 mS/cm in start time 1 but is stable around 9.4 mS/cm for the event in start time 2. This can make analysis using setpoints tricky: some utilities use “seasonal” setpoint values if particular parameters such as temperature or conductivity have entirely different “normal” values at different times of the year or under different operating conditions.

Event start time(s) for a batch are selected during batch creation, described in Section 6.1. Any timestep within the batch’s time period is acceptable.

### ***B.2.5 Simulated Contamination Event Example Calculation***

This section walks through the calculations used to generate a sample simulated contamination event.

Table B-7 shows the baseline data upon which the sample event will be simulated. The date range and polling interval, described in Section B.1, has already been applied to produce this data. For simplicity, only chlorine data will be considered.

**Table B-7. Example Baseline Dataset**

Timestep	CL2
1/12/2012 00:00	0.9
1/12/2012 00:02	0.93
1/12/2012 00:04	0.93
1/12/2012 00:06	0.92
1/12/2012 00:08	0.91
1/12/2012 00:10	0.89
1/12/2012 00:12	0.89
1/12/2012 00:14	0.91
1/12/2012 00:16	0.93
1/12/2012 00:18	0.94
1/12/2012 00:20	0.93
1/12/2012 00:22	0.92
1/12/2012 00:24	0.92
1/12/2012 00:26	0.91
1/12/2012 00:28	0.91
1/12/2012 00:30	0.9

For this example, EDDIES default **Contaminant\_A** will be simulated (as shown in Table B-6, its reaction expressions for chlorine is  $-0.2 \cdot X$ ). The event start time will be **1/12/2012 00:10** and the peak contaminant concentration will be **5 mg/L**. Finally, the sample event profile shown in Table B-8 will be used (the EDDIES default event profiles are too long for this simple example).

**Table B-8. Sample Event Profile**

Timestep	% of Peak Concentration
1	0.1
2	0.25
3	0.5
4	1.0
5	0.75
6	0.5

Table B-9 shows the calculations used to generate the changes in chlorine caused by the simulated contamination event. The event timesteps are bolded and shown in white.

- The percentage of peak concentration for each timestep is obtained by copying the values from the event profile (shown in Table B-8), beginning at the specified event start time (1/12 00:10). All other timesteps have a percentage of 0.
- The contaminant concentration, shown in the fourth column, is calculated by multiplying the percentage of peak concentration for each timestep (column 2) by the peak concentration chosen for the event (5 mg/L).
- The resulting change in chlorine is calculated by applying the contaminant concentration (column 3) to the contaminant reaction factor, which is  $-0.2 \cdot X$  as noted above.

**Table B-9. Chlorine Changes Resulting from Simulated Event**

Timestep	Percentage of Peak Concentration	Resulting Contaminant Concentration	Resulting Chlorine Change
1/12/2012 00:00	0	0	0
1/12/2012 00:02	0	0	0
1/12/2012 00:04	0	0	0
1/12/2012 00:06	0	0	0
1/12/2012 00:08	0	0	0
<b>1/12/2012 00:10</b>	<b>0.1</b>	<b>0.5</b>	<b>-0.1</b>
<b>1/12/2012 00:12</b>	<b>0.25</b>	<b>1.25</b>	<b>-0.25</b>
<b>1/12/2012 00:14</b>	<b>0.5</b>	<b>2.5</b>	<b>-0.5</b>
<b>1/12/2012 00:16</b>	<b>1</b>	<b>5</b>	<b>-1</b>
<b>1/12/2012 00:18</b>	<b>0.75</b>	<b>3.75</b>	<b>-0.75</b>
<b>1/12/2012 00:20</b>	<b>0.5</b>	<b>2.5</b>	<b>-0.5</b>
1/12/2012 00:22	0	0	0
1/12/2012 00:24	0	0	0
1/12/2012 00:26	0	0	0
1/12/2012 00:28	0	0	0
1/12/2012 00:30	0	0	0

Table B-10 shows how these chlorine changes are combined with the baseline data to get the final event chlorine values. First, the baseline chlorine values shown in Table B-7 are repeated, followed by the chlorine changes calculated in Table B-9. These two columns are added together to yield the final event chlorine value. Note that the calculated chlorine value for 00:16 is negative. Since the minimum acceptable value for the chlorine parameter type is zero, this negative chlorine values is overwritten with zero.

**Table B-10. Simulated Event Dataset**

Timestep	Baseline Chlorine	Chlorine Change, X	Event Chlorine
1/12/2012 00:00	0.9	0	0.9
1/12/2012 00:02	0.93	0	0.93
1/12/2012 00:04	0.93	0	0.93
1/12/2012 00:06	0.92	0	0.92
1/12/2012 00:08	0.91	0	0.91
1/12/2012 00:10	0.93	<b>-0.1</b>	0.83
1/12/2012 00:12	0.92	<b>-0.25</b>	0.67
1/12/2012 00:14	0.9	<b>-0.5</b>	0.4
1/12/2012 00:16	0.88	<b>-1</b>	-0.12 -> 0
1/12/2012 00:18	0.91	<b>-0.75</b>	0.16
1/12/2012 00:20	0.93	<b>-0.5</b>	0.43
1/12/2012 00:22	0.92	0	0.92
1/12/2012 00:24	0.92	0	0.92
1/12/2012 00:26	0.91	0	0.91
1/12/2012 00:28	0.91	0	0.91
1/12/2012 00:30	0.9	0	0.9

### B.3 Contaminant and Event Profile Libraries

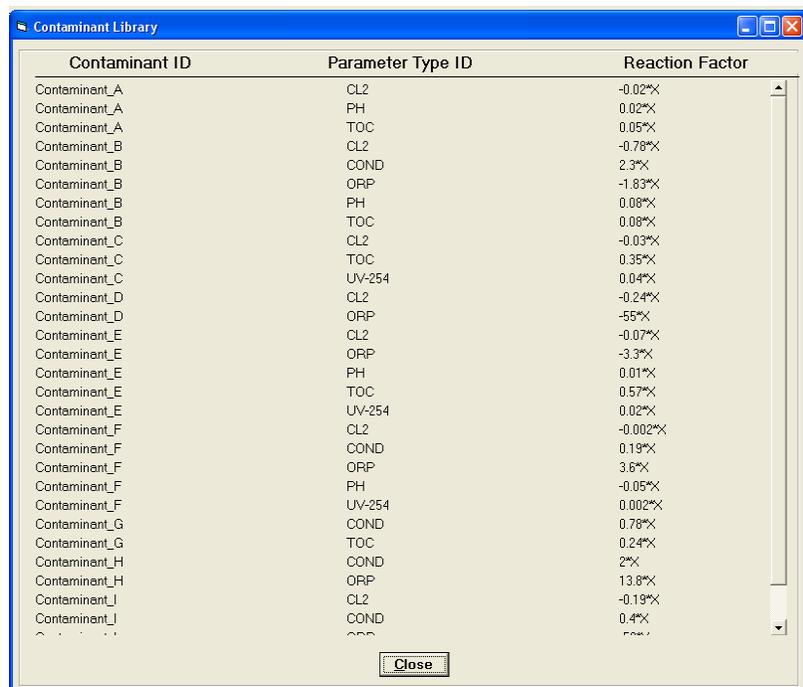
When specifying events during batch creation, the user selects from the contaminants and event profiles currently defined in the EDDIES database. This section describes how to view, add and edit items to these libraries.

### B.3.1. Contaminant Library

This section describes use of the contaminant library.

#### B.3.1.1 Viewing Contaminants

From the 'View' menu, click the 'Contaminant Library' button to view the contaminants in the EDDIES database. There is a row for each parameter and contaminant combination. Figure B-6 shows an example, listing the contaminants in the EDDIES default database. For example, Contaminant\_A is listed three times, each time with a different parameter type and corresponding reaction factor.



Contaminant ID	Parameter Type ID	Reaction Factor
Contaminant_A	CL2	-0.02*X
Contaminant_A	PH	0.02*X
Contaminant_A	TOC	0.05*X
Contaminant_B	CL2	-0.78*X
Contaminant_B	COND	2.3*X
Contaminant_B	ORP	-1.83*X
Contaminant_B	PH	0.08*X
Contaminant_B	TOC	0.08*X
Contaminant_C	CL2	-0.03*X
Contaminant_C	TOC	0.35*X
Contaminant_C	UV-254	0.04*X
Contaminant_D	CL2	-0.24*X
Contaminant_D	ORP	-55*X
Contaminant_E	CL2	-0.07*X
Contaminant_E	ORP	-3.3*X
Contaminant_E	PH	0.01*X
Contaminant_E	TOC	0.57*X
Contaminant_E	UV-254	0.02*X
Contaminant_F	CL2	-0.002*X
Contaminant_F	COND	0.19*X
Contaminant_F	ORP	3.6*X
Contaminant_F	PH	-0.05*X
Contaminant_F	UV-254	0.002*X
Contaminant_G	COND	0.78*X
Contaminant_G	TOC	0.24*X
Contaminant_H	COND	2*X
Contaminant_H	ORP	13.8*X
Contaminant_I	CL2	-0.19*X
Contaminant_I	COND	0.4*X

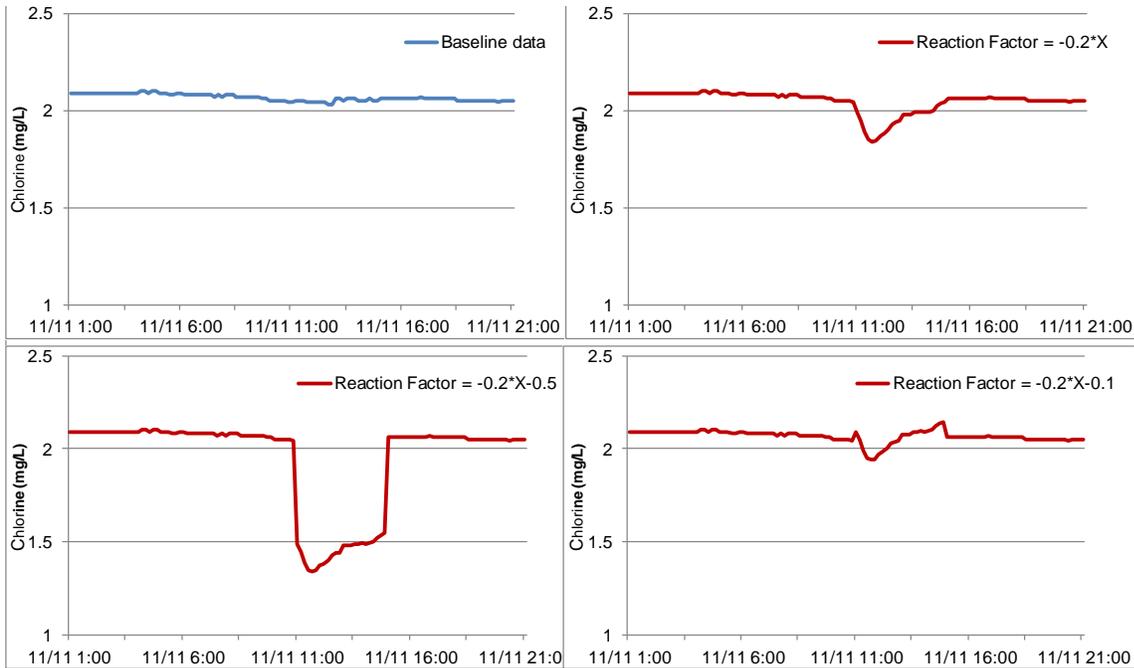
Figure B-6. Contaminant Library Window

#### B.3.1.2 Adding Contaminants

Contaminants can be added using the EDDIES interface or through a CSV import. In either case, a unique contaminant name must be entered, a previously defined parameter type must be specified and a reaction factor must be entered. The reaction factor is entered as an expression where  $X$  is the contaminant concentration, using any string of numbers, mathematical operators and parentheses. EDDIES follows the order of operations when evaluating these factors. '\*' is the symbol for multiplication.

Using linear reaction factors with no constant term is recommended to ensure that the water quality changes are as expected. Figure B-7 shows the unintended changes in chlorine that would result from two different reaction factors containing constant terms. The top plots show the baseline data and a plot using a reaction factor without a constant term ( $-0.2*X$ ), which produces a typical drop in chlorine. The bottom two plots use the same multiple ( $-0.2*X$ ) but add a constant term. The first uses the expression  $-0.2*X - 0.5$ . With the large constant, there is an immediate drop of 0.5 mg/L as soon as the event starts which overpowers the additional changes caused by the changes in concentration. The second plot uses a

reaction factor of  $-0.2 \cdot X - 0.1$ , and at small concentrations this actually causes an increase in chlorine, as can be seen at the end of the event. Use of non-linear reaction factors also can produce undesired behavior.



**Figure B-7. Simulated Event Using Constant Terms in Reaction Factor**

### **Adding Contaminants using the EDDIES Interface**

One contaminant can be added at a time using the EDDIES interface.

1. From the 'Add' menu, click the 'Contaminant' button.
2. In the 'Add a Contaminant' window that pops up, enter a contaminant ID. This must be unique and cannot contain spaces or commas. Click the 'Add' button to save the contaminant ID to the database.
3. Reaction factors are entered for one parameter type at a time on the 'Add Parameter Types and Reaction Factors' window that pops up, shown in Figure B-8.
  - a. Select a parameter type impacted by the contaminant from the dropdown list. Note that the desired parameter type must already be defined (as described in Section 4.1.1).
  - b. Enter the associated reaction factor/expression where  $X$  is the contaminant concentration. Any string of numbers, mathematical operators and parentheses can be used.
  - c. If additional parameter types are impacted by the contaminant, click the 'Add' button. This will bring up a blank 'Add Parameter Types and Reaction Factors' window. Enter the reaction information for the next impacted parameter type.
  - d. Click the 'Done' button when all parameter types impacted by the contaminant have been added to save the contaminant information to the database.

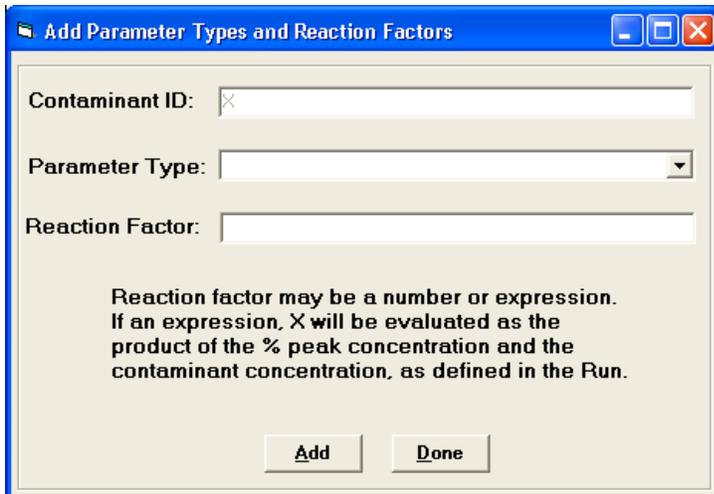


Figure B-8. 'Add Parameters and Reaction Factors' Window

### **Adding Contaminants using a CSV File Import**

Multiple contaminants can be added to the EDDIES database at once by importing a CSV file.

1. Create a CSV file in the format specified in Appendix A.
2. From the 'Import Manager' tab, click the 'Import Event Simulation Properties' button.
3. From the 'Import Analysis Data' menu that pops up, click 'Contaminant Library.'
4. Navigate to the desired CSV file in the 'Select CSV File' browser that pops up and click 'Open.'

If the import is successful, a window showing the number of rows added to the database pops up

#### ***B.3.1.3 Editing Contaminants***

Once contaminants have been added, they cannot be edited or deleted. If revisions are necessary, a new contaminant must be added.

### ***B.3.2. Event Profile Library***

The following sections detail how to view, add and edit event profiles in the EDDIES database.

#### ***B.3.2.1 Viewing Event Profiles***

From the 'View' menu, click the 'Event Profiles' button to view a list of the event profiles in the EDDIES database. There is a row for each event profile and timestep combination. Figure B-9 shows an example. In this example, PROFILE\_C has 23 timesteps and is thus listed 23 times, each time with a different step and associated percentage of peak concentration.

Event Profile	Percent Peak Concentration	Step
PROFILE_A	0.01	1
PROFILE_A	0.02	2
PROFILE_A	0.04	3
PROFILE_A	0.06	4
PROFILE_A	0.09	5
PROFILE_A	0.12	6
PROFILE_A	0.15	7
PROFILE_A	0.18	8
PROFILE_A	0.21	9
PROFILE_A	0.25	10
PROFILE_A	0.3	11
PROFILE_A	0.35	12
PROFILE_A	0.4	13
PROFILE_A	0.46	14
PROFILE_A	0.52	15
PROFILE_A	0.59	16
PROFILE_A	0.65	17
PROFILE_A	0.7	18
PROFILE_A	0.74	19
PROFILE_A	0.78	20
PROFILE_A	0.81	21
PROFILE_A	0.84	22
PROFILE_A	0.87	23
PROFILE_A	0.89	24
PROFILE_A	0.91	25
PROFILE_A	0.93	26
PROFILE_A	0.95	27
PROFILE_A	0.96	28
PROFILE_A	0.97	29

**Figure B-9. View Event Profile Library**

### *B.3.2.2 Adding Event Profiles*

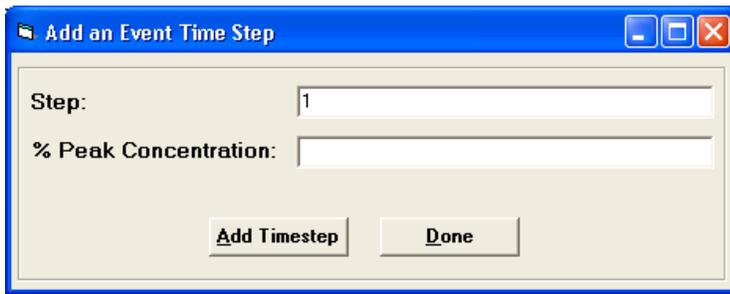
This section details the process of adding event profiles to the EDDIES database. Event profiles can be added using the EDDIES interface or through a CSV import.

It is recommended that several timesteps with a percent peak concentration of zero be added at the end of each event profile. Appendix D describes that only EDS alerts that occur within an event period are considered detections. However, EDSs often produce an alert due to the water quality change that occurs as the contaminant concentration returns from its peak to zero. With these extra timesteps, alerts produced just after the event has ended are captured. All EDDIES default event profiles contain this.

#### **Adding Event Profiles using the EDDIES Interface**

One event profile is added at a time using this method.

1. From the ‘Add’ menu, select the ‘Event Profile’ button.
2. In the ‘Add an Event Profile’ window that pops up, enter an Event Profile ID. This must be unique and contain no spaces or commas. Click ‘Add’ to save the Event Profile ID to the database.
4. One timestep at a time is added on the ‘Add an Event Time Step’ window that pops up, shown in Figure B-10. The timestep field is automatically populated – starting at one and increasing by one for each timestep that is added.
  - a. Enter the desired percentage of peak concentration as a number (not a percentage).
  - b. If additional timesteps are desired, click the ‘Add Timestep’. This will bring up a blank ‘Add an Event Time Step’ window. Enter the percentage of peak concentration for the next timestep. Repeat this step until all event profile timesteps have been added.
  - c. Click the ‘Done’ button when all event profile timesteps have been added to save the event profile information to the database.



**Figure B-10. 'Add an Event Profile Time Step' Window**

### **Adding Event Profiles using a CSV File Import**

Multiple event profiles can be added to the EDDIES database at once by importing a CSV file,

1. Created a CSV file formatted as specified in Appendix A.
2. From the 'Import Manager' tab, click the 'Import Event Simulation Properties' button.
3. From the 'Import Analysis Data' menu that pops up, click the 'Event Profile Library' button.
4. Navigate to the desired CSV file in the browser window that pops up and click the 'Open' button.

If the import is successful, a window showing the number of event profile steps added to the database pops up.

#### ***B.3.2.3 Editing Event Profiles***

Once event profiles have been added, they cannot be edited or deleted. If revisions are necessary, a new event profile must be added.

## Appendix C: Export File Formats

This appendix describes the format of each EDDIES export file type. All export files are in CSV format.

### C.1 'Test Dataset' Export

Test Datasets can be exported from the Launch Manager, as described in Section 6.2. There are two formats available for exporting baseline data: time-series and parameter-based formats. Additionally, EDDIES can export data in both 12-hour and 24-hour time format.

#### C.1.1 Time-series Format

In time-series format, each row includes the values for all parameters for a single timestep. Time-series export files have the following fields in the header row: TIME\_STEP, EVENT\_STATUS, Parameter IDs for all parameters associated with the given location.

The remaining rows in time-series files have the following values:

- Timestep: A string in one of the following formats: 12-hour (MM/DD/YYYY hh:mm am) or 24-hour time format (MM/DD/YYYY hh:mm).
- Event status: -1, 0 or 1. more details on event status can be found in Appendix D.
- Parameter values: The values for the parameters are in the order the IDs were listed in the header row.

Table C-1 shows the contents of an example export file that uses the 24-hour time format.

**Table C-1. Sample Imported Data Export File in Time-series Format**

TIME_STEP	EVENT_STATUS	A_COND	A_CI2	A_PH	A_PRESSURE	A_PUMP1
11/1/2007 0:00	0	45.49	1.57	9.18	35.46	0
11/1/2007 0:05	0	45.15	1.81	9.18	35.46	0
11/1/2007 0:10	0	44.94	1.53	9.17	35.46	0

#### C.1.2 Parameter-based Format

In parameter-based format, there can be multiple rows for each timestep with each row containing data for a single parameter for a single timestep. Parameter-based export files have the following fields in the header row: TIME\_STEP, EVENT\_STATUS, SAMPLE\_VALUE, SAMPLE\_QUALITY, PARAMETER\_ID.

The remaining rows have the following values:

- Timestep: A string in one of the following formats: 12-hour (MM/DD/YYYY hh:mm am) or 24-hour time format (MM/DD/YYYY hh:mm).
- Event status: -1, 0 or 1.
- Sample value: The values for the parameter for that timestep.
- Sample quality: Normal or Bad. This field is optional and not used by EDDIES. Thus, it is often blank.
- Parameter ID: The PARAMETER\_ID of the data reported.

Table C-2 shows an example data export file in parameter-based format that uses the 24-hour time format.

**Table C-2. Sample Imported Data Export File in Parameter-based Format**

TIME_STEP	EVENT_STATUS	SAMPLE_VALUE	SAMPLE_QUALITY	PARAMETER_ID
11/1/2007 0:00	0	45.49	Normal	A_COND
11/1/2007 0:00	0	1.57	Normal	A_CL2
11/1/2007 0:00	0	9.18	Normal	A_PH

## C.2 'Log Results and Data' Export

The 'Log Results and Data' exports are generated as batches are executed within EDDIES, as described in Section 6.2. The following fields are in the header row: TIME\_STEP, Parameter IDs for all parameters associated with the given location, ALERT\_STATUS, ALERT\_LEVEL, ANALYSIS\_COMMENTS, CONTRIBUTING\_PARAMETERS, EVENT\_STATUS.

The remaining rows have the following values. See Section 1.2 for a description of the EDS outputs (ALERT\_STATUS, ALERT\_LEVEL, ANALYSIS\_COMMENTS, CONTRIBUTING\_PARAMETERS).

- Timestep: A string in 24-hour time format (MM/DD/YYYY hh:mm).
- Parameter ID: The PARAMETER\_ID of the data reported.
- Alert status: Typically a value of 0, 1 or 2.
- Alert level: Any real number.
- Analysis comments: an optional, alphanumeric string outputted by the EDS.
- Contributing parameters: an optional, alphanumeric string outputted by the EDS.
- Event status: -1, 0 or 1.

Table C-3 shows the contents of an example Log results and data export file.

**Table C-3. Sample Log Results and Sample Data File**

TIME_STEP	CI2	PH	ALERT_STATUS	ALERT_LEVEL	ANALYSIS_COMMENTS	CONTRIBUTING_PARAMETERS	EVENT_STATUS
11/1/2007 0:00	1.57	7.9	2	0	Insufficient history		0
11/1/2007 0:05	1.81	7.92	0	0			0
11/1/2007 0:10	1.53	7.94	1	0.5		CL2	0

## C.3 'Alerts' Export

The 'Alerts' export file can be exported from the 'Export Manager' tab and contains a list of alerts produced during the selected runs. These use the user-defined alert time window, and the user can choose to have alerts determined using either the Alert level or Alert status. Details on how alerts are determined are provided in Appendix D.

Table C-4 shows an example 'Alerts' export file. The first two rows of the file define the user-specified conditions of the export.

- The first row lists the method used for identifying the alerts: “Analysis by Alert Status” or “Alert Threshold=y”, where y is the user-defined discrimination threshold.
- The second row lists the user-defined alert time window employed when the file was exported.

**Table C-4. Sample ‘Alerts’ Export**

Alert Threshold=0.9				
Alert Time Window=0				
RUN_ID	ALERT_START	ALERT_END	ALERT_COMMENT	
1	10/14/2007 13:00	10/15/2007 11:50	False Positive	
1	10/15/2007 15:20	10/15/2007 15:35	False Positive	
1	10/23/2007 12:15	10/23/2007 12:45	True Positive	Baseline Event Detected
1	12/20/2007 10:50	12/20/2007 10:55	False Positive	
2	11/5/2007 10:10	11/5/2007 12:10	True Positive	Simulated Event Detected
3	12/25/2007 14:50	12/25/2007 16:35	True Positive	Simulated Event Detected

The next row is a fixed header row defining the alert information using the following fields: RUN\_ID, ALERT\_START, ALERT\_END, ALERT\_COMMENT. The remaining rows contain a list of the alerts output based on the user-defined conditions, described with the following metrics:

- Run ID: The numeric identifier of the run in which the alert occurred.
- Alert start: The first timestep of the alert, provided as a string in 12-hour or 24-hour time format.
- Alert end: The last timestep of the alert, provided as string in 12-hour or 24-hour time format.
- Alert comment: Classification of the alert as either a true positive or as a false positive.
  - True Positive: An alert beginning during an event; an alert whose first timestep has EVENT\_STATUS =1. For each true positive, the fifth column (this column has no header) denotes whether the alert detected a baseline event or a simulated contamination event.
    - Baseline Event Detected: An alert beginning during a user-defined baseline event.
    - Simulated Event Detected: An alert during a simulated contamination event.
  - False Positive: An alert beginning during normal baseline data; an alert whose first timestep has EVENT\_STATUS =0. False positive alerts are generated exclusively in baseline runs.

#### C.4 Export Results and Data

EDS results – with or without the corresponding data included – can be exported through the ‘Export Manager’ tab. See Section 1.2 for more details on the EDS outputs fields.

When results with data are exported, the exported results have the following fields in the header row: TIME\_STEP, Parameter IDs for all parameters associated with the given location, ALERT\_STATUS, ALERT\_LEVEL, ANALYSIS\_COMMENTS, CONTRIBUTING\_PARAMETERS, EVENT-STATUS.

The remaining rows have the following values:

- TIME\_STEP: A string in one of the following formats: 12-hour (MM/DD/YYYY hh:mm am) or 24-hour time format (MM/DD/YYYY hh:mm).
- Parameter values: Values for the parameters are in the order they were listed in the header row.

- ALERT STATUS: typically a value of 0 or 1.
- ALERT LEVEL: Any value.
- ANALYSIS\_COMMENTS: Any alphanumeric string (max length = 1000 characters). Optional.
- CONTRIBUTING PARAMETERS: An optional, alphanumeric string. Optional.
- EVENT\_STATUS: -1, 0 or 1.

Table C-5 shows an example Results and Data export file.

**Table C-5. Sample Results with Sample Data Export File**

TIME_STEP	A_CL2	A_CON D	ALERT_ STATU S	ALERT_ LEVEL	ANALYSIS_ COMMENTS	CONTRIBUTIN G_PARAMETE RS	EVENT_ STATU S
11/2/2007 0:00	2.2	44.98	0	0	Filling window		0
11/2/2007 0:05	2.2	44.99	0	0			0
11/2/2007 0:10	2.3	44.98	0	0			0
11/2/2007 0:15	1.9	44.97	0	0.2		CL2	0
11/2/2007 0:20	1.8	44.98	1	0.7		CL2	0

When results are exported without data, the exported results have the following fields in the header row: TIME\_STEP, ANALYSIS\_COMMENTS, ALERT\_STATUS, ALERT\_LEVEL, CONTRIBUTING\_PARAMETERS, EVENT\_STATUS. The format of these fields are the same as described above in the case that data is included.

A file without data included would look identical to Table C-5 except the columns showing data (the second and third columns) would be missing: ALERT\_STATUS would immediately follow the TIME\_STEP column.

## C.5 Analysis Data Export

EDDIES contains functionality to export a detailed or basic summary of the EDS' performance on user-selected run(s). All included metrics are described in detail in Appendix D.

The first several rows of both the detailed and basic files show the batch and analysis variable settings used during file creation. The remaining rows have one of two formats. For metrics that do not depend on the discrimination threshold, the metric name is in the first column, followed by the corresponding value in the second. For metrics that do depend on the threshold, the first column gives the metric name and the remaining column(s) give the performance value for a particular discrimination threshold setting.

Table C-6 gives an example basic export file and Table C-7 shows a detailed file based on the same runs.

**Table C-6. Sample Basic Analysis Data Export File**

Number of Simulated Event Runs in Analysis	96				
Number of Batches in Analysis	1				
Polling Interval	5				
---ANALYSIS PARAMETER SETTINGS---					
Discrimination Threshold Increment	0.25				
Discrimination Threshold Minimum	0				
Discrimination Threshold Maximum	1				
Alert Time Window	5				
Use Net Response for Contamination Event Analysis	no				
Required Ratio for Detection	0.001				
---SUMMARY OF ANALYSIS DATA---					
Number of Timesteps Identified as Normal by User in Uploaded Data	68055				
Number of Timesteps Identified as Bad Data by User in Uploaded Data	0				
Number of Abnormal Events Identified by User in Uploaded Data	4				
Number of Simulated Events	96				
DISCRIMINATION THRESHOLD					
	0	0.25	0.5	0.75	1
***** OVERALL EDS PERFORMANCE *****					
Percent of Normal Timesteps that are False Positives	100%	1.01%	0.87%	0.77%	0.41%
Percent of Events Detected - User-Identified and Simulated Events	100%	74%	74%	74%	69%
Median Time to Detect for All Events - User-Identified and Simulated Events (timesteps)	0	7	8	9	16
***** A LITTLE MORE DETAIL *****					
---EDS Performance on Uploaded Data Identified by User as Normal---					
Number of Invalid Alerts	1	53	46	42	33
Average Invalid Alert Length (timesteps)	68256	12.9434	12.8261	12.4286	8.4848
---EDS Performance on Uploaded Data Identified by User as Abnormal---					
Number of Alerts on Abnormal Data (Valid Alerts)	0	3	3	3	2
Percent of User-Identified Events Detected	100%	75%	75%	75%	50%
Average Percent of Event Timesteps the EDS Alerts On for Detected User-Identified Events	100%	34.20%	29.95%	25.70%	16.17%
Average Time to Detect for Detected User-Identified Events (timesteps)	0	3.33333	4	4.66667	12
Median Time to Detect for All User-Identified Events (timesteps)	0	5.5	6.5	7.5	ND
---EDS Performance on Simulated Events ---					
Percent of Simulated Events Detected	100%	73.96%	73.96%	73.96%	69.79%
Average Percent of Event Timesteps the EDS Alerts On for Detected Simulated Events	100%	49.79%	46.53%	43.27%	29.10%
Average Time to Detect for Detected Simulated Events (timesteps)	0	7.85915	8.85915	9.85915	14.56716
Median Time to Detect for All Simulated Events (timesteps)	0	8	9	10	16

**Table C-7. Sample Basic Analysis Data Export File**

Number of Simulated Event Runs in Analysis	96				
Number of Batches in Analysis	1				
Polling Interval	5				
---ANALYSIS PARAMETER SETTINGS---					
Discrimination Threshold Increment	0.25				
Discrimination Threshold Minimum	0				
Discrimination Threshold Maximum	1				
Alert Time Window	5				
Use Net Response for Contamination Event Analysis	no				
Required Ratio for Detection	0.001				
***** OVERALL EDS PERFORMANCE *****					
DISCRIMINATION THRESHOLD	0	0.25	0.5	0.75	1
Percent of Normal Timesteps that are False Positives	100%	1.01%	0.87%	0.77%	0.41%
Percent of Events Detected - User-Identified and Simulated Events	100%	74%	74%	74%	69%
Median Time to Detect for All Events - User-Identified and Simulated Events (timesteps)	0	7	8	9	16
Average Time to Detect for Detected Events - User-Identified and Simulated Events (timesteps)	0	7.74324	8.74324	9.74324	14.4928
***** EDS PERFORMANCE ON UPLOADED DATA CLASSIFIED AS NORMAL BY USER *****					
Number of Timesteps Identified as Normal by User in Uploaded Data	68055				
Average Alert Level on Normal Timesteps	9.91E-03				
Minimum Alert Level on Normal Timesteps	0				
First Quartile Alert Level on Normal Timesteps	0.000244				
Median Alert Level on Normal Timesteps	0.000244				
Third Quartile Alert Level on Normal Timesteps	0.000244				
Maximum Alert Level on Normal Timesteps	1				
Standard Deviation of Alert Level on Normal Timesteps	0.09				
---Threshold - Dependent Metrics---					
DISCRIMINATION THRESHOLD	0	0.25	0.5	0.75	1
Number of False Positive Timesteps - EDS incorrectly Alerting for normal timestep	68055	686	590	522	280
Number of True Negative Timesteps - EDS correctly not Alerting for normal timestep	0	67369	67465	67533	67775
Percent of Normal Timesteps that are False Positives	100%	1.01%	0.87%	0.77%	0.41%
Number of Invalid Alerts	1	53	46	42	33
Invalid Alert Frequency (average number of timesteps between Invalid Alerts)	68055	1284.06	1479.46	1620.36	2062.27
Average Invalid Alert Length (timesteps)	68256	12.9434	12.8261	12.4286	8.4848
Minimum Invalid Alert Length (timesteps)	68256	1	1	1	2
First Quartile Invalid Alert Length (timesteps)	68256	8	9.25	10	8
Median Invalid Alert Length (timesteps)	68256	17	16	15	10
Third Quartile Invalid Alert Length (timesteps)	68256	17	16	15	10
Maximum Invalid Alert Length (timesteps)	68256	21	19	17	10
***** EDS PERFORMANCE ON UPLOADED DATA CLASSIFIED AS ABNORMAL BY USER *****					
Number of Abnormal Events Identified by User in Uploaded Data	4				
Total Number of Timesteps in User-Identified Events	201				
Average Alert Level for User-Identified Events	0.196232				
Minimum Alert Level for User-Identified Events	0				
First Quartile Alert Level for User-Identified Events	0.000244				
Median Alert Level for User-Identified Events	0.000244				
Third Quartile Alert Level for User-Identified Events	0.072998				
Maximum Alert Level for User-Identified Events	1				
Standard Deviation Alert Level for User-Identified Events	0.37				
---Threshold - Dependent Metrics---					

Number of Simulated Event Runs in Analysis	96				
DISCRIMINATION THRESHOLD	0	0.25	0.5	0.75	1
Number of Alerts on Abnormal Data (Valid Alerts)	0	3	3	3	2
Number of User-Identified Events Detected (True Positives)	4	3	3	3	2
Number of User-Identified Events Not Detected (False Negatives)	0	1	1	1	2
Percent of User-Identified Events Detected	100%	75%	75%	75%	50%
Number of Detected Timesteps in User-Identified Events	201	44	40	36	20
Average Time to Detect for Detected User-Identified Events (timesteps)	0	3.33333	4	4.66667	12
Minimum Time to Detect for All User-Identified Events (timesteps)	0	4	5	6	11
First Quartile Time to Detect for All User-Identified Events (timesteps)	0	4.75	5.75	6.75	12.5
Median Time to Detect for All User-Identified Events (timesteps)	0	5.5	6.5	7.5	ND
Third Quartile Time to Detect for All User-Identified Events (timesteps)	0	250005	250005	250006	ND
Maximum Time to Detect for All User-Identified Events (timesteps)	0	ND	ND	ND	ND
Average Percent of Event Timesteps the EDS Alerts On for Detected User-Identified Events	100%	34.20%	29.95%	25.70%	16.17%
Minimum Percent of Event Timesteps the EDS Alerts On for All User-Identified Events	100%	0%	0%	0%	0%
First Quartile Percent of Event Timesteps the EDS Alerts On for All User-Identified Events	100%	15.74%	14.81%	13.89%	0%
Median Percent of Event Timesteps the EDS Alerts On for All User-Identified Events	100%	27.49%	25.88%	23.54%	6.17%
Third Quartile Percent of Event Timesteps the EDS Alerts On for All User-Identified Events	100%	37.40%	33.52%	28.93%	14.26%
Maximum Percent of Event Timesteps the EDS Alerts On for All User-Identified Events	100%	47.62%	38.10%	30%	20%
***** EDS PERFORMANCE ON UPLOADED DATA CLASSIFIED AS BAD QUALITY DATA BY USER *****					
Number of Timesteps Identified as Bad Data by User in Uploaded Data	0				
---Threshold - Dependent Metrics---					
DISCRIMINATION THRESHOLD	0	0.25	0.5	0.75	1
Number of Timesteps with Bad Quality Data that the EDS Alerts on	NA	NA	NA	NA	NA
Number of Alerts on Bad Quality Data	NA	NA	NA	NA	NA
***** EDS PERFORMANCE ON SIMULATED CONTAMINATION EVENTS *****					
Number of Simulated Events	96				
Total Number of Timesteps in Simulated Events	4848				
Average Net Response for Simulated Events	0.275271				
Minimum Net Response for Simulated Events	-0.00293				
First Quartile Net Response for Simulated Events	0				
Median Net Response for Simulated Events	0				
Third Quartile Net Response for Simulated Events	0.805908				
Maximum Net Response for Simulated Events	0.999756				
Trigger Accuracy	0.87				
---Threshold - Dependent Metrics---					
DISCRIMINATION THRESHOLD	0	0.25	0.5	0.75	1
Number of True Positives - Simulated Events Detected	96	71	71	71	67
Number of False Negatives - Simulated Events Not Detected	0	25	25	25	29
Percent of Simulated Events Detected	100%	73.96%	73.96%	73.96%	69.79%
Number of Detected Timesteps for Simulated Events	4848	1430	1334	1238	779
Average Time to Detect for Detected Simulated Events (timesteps)	0	7.85915	8.85915	9.85915	14.5672
Minimum Time to Detect for All Simulated Events (timesteps)	0	5	6	7	12
First Quartile Time to Detect for All Simulated Events (timesteps)	0	5	6	7	12
Median Time to Detect for All Simulated Events (timesteps)	0	8	9	10	16
Third Time to Detect for All Simulated Events (timesteps)	0	ND	ND	ND	ND
Maximum Time to Detect for All Simulated Events (timesteps)	0	ND	ND	ND	ND
Average Percent of Event Timesteps the EDS Alerts On for Detected Simulated Events	100%	49.79%	46.53%	43.27%	29.10%
Minimum Percent of Event Timesteps the EDS Alerts On for All Simulated Events	100%	0%	0%	0%	0%

Number of Simulated Event Runs in Analysis	96				
First Quartile Percent of Event Timesteps the EDS Alerts On for All Simulated Events	100%	0%	0%	0%	0%
Median Percent of Event Timesteps the EDS Alerts On for All Simulated Events	100%	47.89%	45.07%	42.25%	28.17%
Third Quartile Percent of Event Timesteps the EDS Alerts On for All Simulated Events	100%	56.67%	53.33%	50%	33.33%
Maximum Percent of Event Timesteps the EDS Alerts On for All Simulated Events	100%	56.67%	53.33%	50%	33.33%

## C.6 Setpoint Sensitivity Analysis Export

The setpoint sensitivity analysis export is executed from the ‘Export Manager’ tab. The user has the option to perform the export and calculate the metrics using either a low or high setpoint threshold. Section E.2 discusses interpretation of these files.

Table C-8 shows the contents of an example setpoint sensitivity analysis export file, which is very similar to the basic analysis export file. This example uses a low setpoint threshold, though the format for a high setpoint sensitivity analysis export is the same.

**Table C-8. Sample Setpoint Low Export File**

Number of Simulated Event Runs in Analysis	5					
Number of Batches in Analysis	1					
Polling Interval	5					
Setpoint Type	High to detect increase in parameter value					
---ANALYSIS PARAMETER SETTINGS---						
Discrimination Threshold Increment	0.5					
Discrimination Threshold Minimum	0.5					
Discrimination Threshold Maximum	3					
Alert Time Window	0					
Use Net Response for Contamination Event Analysis	no					
Required Ratio for Detection	0.1					
---SUMMARY OF ANALYSIS DATA---						
Number of Timesteps Identified as Normal by User in Uploaded Data	299777					
Number of Timesteps Identified as Bad Data by User in Uploaded Data	0					
Number of Abnormal Events Identified by User in Uploaded Data	6					
Number of Simulated Events	5					
SETPOINT VALUE	0.5	1	1.5	2	2.5	3
***** OVERALL EDS PERFORMANCE *****						
Percent of Normal Timesteps that are False Positives	97.8%	90.2%	73.4%	50.9%	22.4%	22.3%
Number of Invalid Alerts	146	627	917	1026	23	20
Average Invalid Alert Length (timesteps)	2011.3	431.7	240.3	148.9	2929.9	3363.6
Percent of Events Detected - User-Identified and Simulated Events	100%	100%	100%	100%	90.9%	90.9%
Percent of User-Identified Events Detected	100%	100%	100%	100%	83.3%	83.3%
Percent of Simulated Events Detected	100%	100%	100%	100%	100%	100%
Median Time to Detect for All Events - User-Identified and Simulated Events (timesteps)	0	0	0	0	0	0

## Appendix D: Key Terms and Analysis Methodology

This appendix provides an in-depth explanation of the analyses performed by EDDIES in the analysis exports. Section D.1 describes the output generated by the EDSs. Section D.2 defines the key terms and performance metrics used by EDDIES and Section D.3 shows example calculations for all metrics.

Section 7 describes how to generate these export files and Appendix C gives the format of each file.

### D.1 EDS Output

EDSs produce three output values which EDDIES uses to calculate the various performance metrics. As noted below, the *trigger parameters* output is not outputted by all EDSs.

- **Alert level** (required): A real number reflecting how certain the EDS is that conditions are anomalous, with higher values indicating more certainty that a water quality anomaly is occurring. This measure was originally called event probability as it was practically interpreted to be the EDS's assessment of how likely it is that an event is occurring. This term was changed because the maximum alert level varies by EDS: many output values greater than 1. The alert level output is used in the 'Analysis Data' export and the 'Alerts' export if the user chooses.
- **Alert status** (required): A binary normal/abnormal indication which precisely identifies when the EDS is alerting. `ALERT_STATUS = 1` indicates that the EDS is alerting due to an assumed anomaly. The alert status output is used in the 'Alerts' export if the user chooses.
- **Trigger parameters** (optional): During periods of elevated alert level, a list of the water quality parameter(s) whose values caused the increase. This output is used in the 'Analysis Data' export if the user chooses.

In general, the alert level and alert status are directly related and alerting is based on an internal *discrimination threshold*. The discrimination threshold is generally a configurable variable for EDSs.

- **Discrimination threshold**: A real value which discriminates between normal and abnormal EDS outputs. A timestep is an alerting timestep if the *alert level*  $\geq$  the *discrimination threshold*.

Note that if this relationship between alert level and alert status does not apply for an EDS under evaluation, exports using the alert level *are not valid*.

To illustrate this relationship, Figure D-1 shows water quality data and EDS output for a sample event impacting chlorine. For the EDS shown, the discrimination threshold was set to one and thus produced the alert status values shown in red: there was one alert lasting 2.8 hours. The purple and green points show how alerting would change if the discrimination threshold were changed. If the threshold were lowered to 0.5 there would be two alerts, one lasting two hours and the second lasting 5 hours (this second alert corresponds to the alert produced with the threshold of 1 but is longer). If the discrimination threshold were raised to 1.5, no alerts would be produced during the period shown.

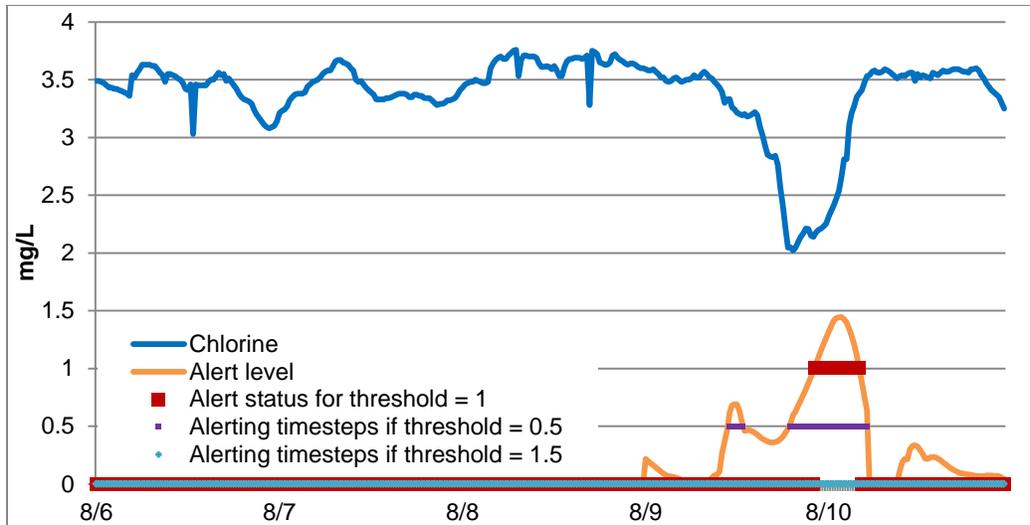


Figure D-1. Example EDS Output

The ‘Analysis Data’ and ‘Sensitivity Analysis’ exports use the *alert level*, presenting what performance would be over a range of *discrimination thresholds*. In the ‘Analysis Data’ export, the user may also choose to include a measure of the accuracy of the *trigger parameters* output. For the ‘Alerts’ export, the user selects whether performance should be calculated using the *alert status* or using a single user-defined *discrimination threshold*.

## D.2 Key Terms and Metrics

This section defines key terms and the performance metrics included in the ‘Analysis Data’ and ‘Sensitivity Analysis’ export files. Section D.3 shows example calculations for all metrics.

### D.2.1 Key Terms

The following key terms are used throughout EDDIES.

#### Datasets

- **Imported data:** Utility data uploaded into EDDIES by the user.
- **Baseline Dataset:** A dataset containing only imported data, generated as described in Appendix B. Baseline datasets are used to calculate invalid alert and baseline event detection metrics.
- **Simulated Event Dataset:** A dataset containing a contamination event generated by EDDIES based on the baseline dataset and event properties selected during batch creation. Appendix B describes how simulated datasets are generated. Event detection is analyzed using these datasets.
- **Run:** Defined by a dataset and the EDS configuration used. There are baseline runs (in which a baseline dataset is presented to the EDS) and event runs (which use simulated event datasets).
- **Batch:** A set of runs that are implemented together. A batch consists of one baseline dataset and any number of simulated event datasets specified by the user. All runs in a batch are based on the same EDS configuration, monitoring station and baseline dataset.

## Timestep Classification

- **Event status:** A user-specified value for each timestep indicating the status of data.
  - Event Status = 1: Indicates a baseline event, which is a period of unusual water quality in the utility data. Common causes of abnormal data include unusual system operations, a change in the treatment process, main breaks and nitrification events. This value should be assigned if the user believes the EDS should alert for the timestep.
  - Event Status = -1: Indicates bad quality data. Common causes of inaccurate data include sensor malfunction, station calibration and communications failure. During analysis, output from these timesteps is considered separately. This separate consideration allows utilities to use the results as desired: some utilities have reported that alerts on bad quality data are desirable whereas others would prefer not receiving them but find them straightforward to rule out.
  - Event Status = 0: Indicates normal data. The EDS should not alert for normal timesteps.
- **Baseline Timestep:** A timestep in the baseline dataset. Each baseline timestep is further classified as an event, bad quality data or normal timestep by the user.
- **Event Timestep:** A timestep where EVENT\_STATUS = 1. This can occur within a baseline dataset (where the user specifies the EVENT\_STATUS = 1) or simulated event dataset (where EDDIES assigns the EVENT\_STATUS = 1). EDSs *should alert* on event timesteps.
- **Bad Quality Data Timestep:** A timestep in a baseline dataset with a user-specified EVENT\_STATUS = -1.
- **Normal Timestep:** A baseline timestep which the user does not consider to be an event or bad quality data; a baseline timestep with a user-specified EVENT\_STATUS = 0. Theoretically, EDSs *should not alert* on normal timesteps.
- **Baseline Event:** A user-identified period in the baseline dataset during which an EDS alert is expected; a continuous series of baseline timesteps.
- **Simulated Event:** A continuous period in an event dataset in which EDDIES superimposes a contamination event on imported data.
- **Bad Quality Data Event:** A user-identified period in the baseline dataset where the data is inaccurate; a continuous series of bad quality timesteps.

## Alerting for Timesteps

- **Net Response:** The difference between the alert level generated for a simulated event timestep and the alert level for the corresponding timestep in the baseline dataset. The net response metric is intended to reflect the portion of the alert level directly attributable to the modified water quality in the simulated event, assuming the EDS produces the same alert levels when presented with the same data multiple times.
- **Alerting Timestep:** A timestep for which the EDS is alerting. Depending on the type of analysis; a timestep with ALERT\_STATUS = 1, a timestep with ALERT\_LEVEL  $\geq$  the specified discrimination threshold, a simulated event timestep where the net response  $\geq$  the specified discrimination threshold, a timestep where the water quality value  $\geq$  the parameter's specified setpoint value.
- **False Positive Timestep:** A normal timestep for which the EDS incorrectly alerts; a timestep that is both a normal timestep and an alerting timestep.

- **True Negative Timestep:** A normal timestep for which the EDS correctly does not alert; a timestep that is a normal timestep and not an alerting timestep.

### Alerts

- **Alert:** A single notification of a potential water quality anomaly; a contiguous sequence of timesteps such that there are no more consecutive non-alerting timesteps between alerting timesteps in the sequence than the number of timesteps in the user-defined alert time window size.
- **Valid Alert:** An alert beginning on an event timestep; an alert whose first timestep has EVENT\_STATUS = 1. Valid alerts can occur during baseline events or simulated events. There can be multiple valid alerts during a single event.
- **Invalid Alert:** An alert beginning on a normal timestep; an alert whose first timestep has EVENT\_STATUS = 0. Invalid alerts are only captured in baseline datasets.
- **Bad Quality Data Alert:** An alert beginning on a bad quality data timestep; an alert whose first timestep has EVENT\_STATUS = -1. Bad quality alerts are captured only in baseline datasets.
- **Alert Length:** The number of timesteps over which an alert occurs, including any non-alerting timesteps within the alert as a consequence of the user-defined alert time window; the alert end time minus the alert start time, in timesteps.

### Detections

- **Detected Event:** A baseline or simulated event for which  $\frac{\# \text{ alerting timesteps}}{\# \text{ event timesteps}} \geq \text{required ratio}$ , defined by the user.
- **Missed Events:** An event that is not a detected event.
- **Detected Timestep:** An alerting timestep occurring during a detected event. If an alerting timestep occurs during an event but the required ratio is not met (the event is not detected), the timestep is not considered to be a detected timestep.
- **Time to Detect:** The number of event timesteps that elapse before the first alerting timestep occurs. EDDIES assigns a time to detect of ‘ND’ to undetected events.

#### ***D.2.2 Data Analysis Metrics***

This section describes how each metric in the detailed ‘Analysis Data’ and ‘Sensitivity Analysis’ export files is calculated. Italicized terms in this section are defined in Section D.2.1.

Standard statistical metrics, including quartile and standard deviation calculations, are used repeatedly in the ‘Analysis Data’ file with the intent of giving the user an understanding of the range and “normal” values. These metrics are calculated using the methodology followed by Microsoft Excel.

All performance metrics with units of time are reported in terms of number of timesteps. These metrics can be translated to units of time, such as hours, by dividing number of timesteps by the number of timesteps in the desired unit of time.

### D.2.2.1 Overall EDS Performance Metrics

The ‘Analysis Data’ export file provides the following overall performance metrics.

- **Percent of Normal Timesteps that are False Positives:** The number of *false positive timesteps* divided by the number of *normal timesteps* across all selected baseline runs.
- **Percent of Events Detected - User-Identified and Simulated Events:** The total number of *detected events* in the selected runs (including both *baseline events* and *simulated events*) divided by the total number of events (both *baseline events* and *simulated events*).
- **Median Time to Detect for All Events - User-Identified and Simulated Events (timesteps):** The median *time to detect* for all detected events. This calculation includes all events - even those that were not detected. Undetected are assigned a time to detect of 'ND', and thus if less than half of the events were detected, the median time to detect is given as 'ND'.
- **Average Time to Detect for Detected Events - User-Identified and Simulated Events (timesteps):** The mean *time to detect* for *detected events*. Unlike the median time to detect, *missed events* are not included in this calculation.

### D.2.2.2 EDS Performance on Imported Data Classified as Normal By User

This section describes the performance metrics calculated on *normal timesteps*. These metrics consider only normal timesteps in the baseline dataset: non-event timesteps in event datasets are not considered in analyses.

- **Number of Timesteps Identified as Normal by User in Imported data:** The total number of *normal timesteps* in the selected baseline dataset(s).
- Statistical summary of *alert level* for all *normal timesteps*.
  - **Average Alert Level on Normal Timesteps**
  - **Minimum Alert Level on Normal Timesteps**
  - **First Quartile Alert Level on Normal Timesteps**
  - **Median Alert Level on Normal Timesteps**
  - **Third Quartile Alert Level on Normal Timesteps**
  - **Maximum Alert Level on Normal Timesteps**
  - **Standard Deviation of Alert Level on Normal Timesteps**

The following metrics are calculated for each *discrimination threshold*.

- **Number of False Positive Timesteps - EDS incorrectly Alerting for normal timestep:** The number of *false positive timesteps* in the selected baseline dataset(s).
- **Number of True Negative Timesteps - EDS correctly not Alerting for normal timestep:** The number of the *true negative timesteps*. Note that the sum of the number of false positive timesteps and true negative timesteps equals the number of normal timesteps.
- **Percent of Normal Timesteps that are False Positives:** The number of *false positive timesteps* divided by the number of *normal timesteps*. This is a standard statistical metric also referred to as ‘1-Specificity.’

- **Number of Invalid Alerts:** The number of *invalid alerts* in the selected baseline dataset(s).
- **Invalid Alert Frequency (average number of timesteps between Invalid Alerts):** The number of *normal timesteps* divided by the number of *invalid alerts*.
- Statistical summary of *invalid alert length*, reported in terms of timesteps.
  - **Average Invalid Alert Length (timesteps)**
  - **Minimum Invalid Alert Length (timesteps)**
  - **First Quartile Invalid Alert Length (timesteps)**
  - **Median Invalid Alert Length (timesteps)**
  - **Third Quartile Invalid Alert Length (timesteps)**
  - **Maximum Invalid Alert Length (timesteps)**

### D.2.2.3 EDS Performance for Baseline Events

This section describes the performance metrics calculated for baseline event periods identified as abnormal by the user. These metrics only consider *event timesteps* during *baseline events*.

- **Number of Abnormal Events Identified by User in Imported data:** The number of *baseline events* in the selected baseline dataset(s).
- **Total Number of Timesteps in Baseline Events:** The total number of timesteps in selected *baseline events*.
- Statistical summary of *alert level* for *event timesteps* in the selected *baseline dataset(s)*.
  - **Average Alert Level for Baseline Events**
  - **Minimum Alert Level for Baseline Events**
  - **First Quartile Alert Level for Baseline Events**
  - **Median Alert Level for Baseline Events**
  - **Third Quartile Alert Level for Baseline Events**
  - **Maximum Alert Level for Baseline Events**
  - **Standard Deviation Alert Level for Baseline Events**

The following metrics are calculated for each *discrimination threshold*.

- **Number of Alerts on Abnormal Data (Valid Alerts):** The number of *valid alerts* during *baseline event(s)* in the selected baseline dataset(s).
- **Number of Baseline Events Detected (True Positives):** The number of *detected baseline events*.
- **Number of Baseline Events Not Detected (False Negatives):** The number of *missed baseline events*.
- **Percent of Baseline Events Detected:** The number of *detected baseline events* divided by the total number of *baseline events*.

- **Number of Detected Timesteps in Baseline Events:** The number of *detected timesteps* in the detected *baseline events*.
- Statistical summary of *time to detect* for *baseline events* in the selected baseline dataset(s), in units of timesteps. Except for the average time, these calculations include all events, including *missed events* which are assigned a time to detect of 'ND'.
  - **Average Time to Detect for Detected Baseline Events (timesteps):** *Missed events* are not included in this calculation.
  - **Minimum Time to Detect for All Baseline Events (timesteps):** If the minimum time to detect is 'ND', then none of the events were detected.
  - **First Quartile Time to Detect for All Baseline Events (timesteps):** If the first quartile time to detect is 'ND', then less than 25% of the events were detected.
  - **Median Time to Detect for All Baseline Events (timesteps):** If the median time to detect is 'ND', then less than half of the events were detected.
  - **Third Quartile Time to Detect for All Baseline Events (timesteps):** If the third quartile time to detect is 'ND', then less than 75% of the events were detected.
  - **Maximum Time to Detect for All Baseline Events (timesteps):** If the maximum time to detect is 'ND', then at least one of the events was not detected.
- Statistical summary of percentage of *alarming timesteps* for *baseline events*. For each baseline event, the number of *alarming timesteps* is divided by the total number of *baseline event timesteps* to get this percentage. Thus, the number of values included in these calculations is the number of baseline events. Except for the average percentage which includes only detected events, these statistical calculations include all events. All *missed events* have 0% of timesteps alerting.
  - **Average Percent of Event Timesteps the EDS Alerts On for Detected Baseline Events**
  - **Minimum Percent of Event Timesteps the EDS Alerts On for All Baseline Events**
  - **First Quartile Percent of Event Timesteps the EDS Alerts On for All Baseline Events**
  - **Median Percent of Event Timesteps the EDS Alerts On for All Baseline Events**
  - **Third Quartile Percent of Event Timesteps the EDS Alerts On for All Baseline Events**
  - **Maximum Percent of Event Timesteps the EDS Alerts On for All Baseline Events**

#### D.2.2.4 EDS Performance on Imported data Classified as Bad Quality by User

This section describes the performance metrics calculated on *bad quality data timesteps* in the baseline dataset provided in the detailed 'Analysis Data' export file.

- **Number of Timesteps Identified as Bad Data by User in Imported data:** The number of *bad quality data timesteps* identified by the user in the selected baseline dataset(s).

The following metrics are calculated for each *discrimination threshold*.

- **Number of Timesteps with Bad Quality Data that the EDS Alerts on:** The number of timesteps in the selected baseline dataset(s) that are both *bad quality data timesteps* and *alerting timesteps*.
- **Number of Alerts on Bad Quality Data:** The number of *bad quality data alerts*.

#### D.2.2.5 EDS Performance on Simulated Contamination Events

This section describes the performance metrics calculated on event timesteps in the selected *simulated event dataset(s)*.

- **Number of Simulated Events:** The number of *simulated events*. This is same as the number of simulated event runs selected by the user to export.
- **Total Number of Timesteps in Simulated Events:** The number of timesteps in the *simulated events*.
- Statistical summary of *net response*. Net response is calculated for all *simulated event timesteps* – both alerting and non-alerting.
  - **Average Net Response for Simulated Events**
  - **Minimum Net Response for Simulated Events**
  - **First Quartile Net Response for Simulated Events**
  - **Median Net Response for Simulated Events**
  - **Third Quartile Net Response for Simulated Events**
  - **Maximum Net Response for Simulated Events**
- **Average Trigger Accuracy:** The average trigger accuracy, as described below, across all *detected simulated events*. NA is outputted if the EDS did not output trigger parameters.

Trigger accuracy is calculated for each detected simulated event and is the percentage of parameter types modified by EDDIES when generating the event that are identified as *trigger parameters* by the EDS. A parameter type is considered to be identified if the EDS outputs it as a *trigger parameter* for any *detected timestep* of the event. Equation D.1 is used to calculate the accuracy for a given parameter,  $p$ .

$Accuracy_p = \begin{cases} 1 & \text{if the EDS outputs the parameter as a trigger parameter for any timestep of the event} \\ 0 & \text{if the EDS does not identify the modified parameter for any timestep of the event} \end{cases}$
---

**Equation D.1**

The trigger accuracy for a *simulated event* combines the accuracy of all impacted parameter types, as shown in Equation D.2. An *impacted parameter type* is one modified by EDDIES during generation of the event, which requires that a parameter with the given type be included in the baseline dataset and that the contaminant chosen impacts that parameter type.

*Trigger Accuracy* =

$$\frac{\sum_{p=1}^x \text{Accuracy}_p}{x}$$

where  $x$  is the number of impacted parameter types and  $p$  represents those parameters

#### Equation D.2

Trigger accuracy is not impacted if an EDS outputs an additional trigger parameter – one that is not an impacted parameter type in the given simulated event. See Section D.3 for an example calculation.

The following metrics are calculated for each *discrimination threshold*.

- **Number of True Positives - Simulated Events Detected:** The number of *detected events* in the selected *simulated event datasets*.
- **Number of False Negatives - Simulated Events Not Detected:** The number of *missed events* in the selected *simulated event datasets*.
- **Percent of Simulated Events Detected:** The number of *detected events* in the selected *simulated event datasets* divided by the total number of selected *simulated events*.
- **Number of Detected Timesteps for Simulated Events:** The number of *detected timesteps* in the selected *simulated events*.
- Statistical summary of *time to detect* for the selected *simulated events*, given in units of timesteps. Except for the average time, these calculations include all events, including *missed events* which are assigned a time to detect of 'ND'. See Section D.2.2.3 for more details on what 'ND' in these results means.
  - **Average Time to Detect for Detected Simulated Events (timesteps)**
  - **Minimum Time to Detect for All Simulated Events (timesteps)**
  - **First Quartile Time to Detect for All Simulated Events (timesteps)**
  - **Median Time to Detect for All Simulated Events (timesteps)**
  - **Third Quartile Time to Detect for All Simulated Events (timesteps)**
  - **Maximum Time to Detect for All Simulated Events (timesteps)**
- Statistical summary of percentage of *alerting timesteps* for the selected *simulated events*. For each event, the number of alerting timesteps is divided by the total number of event timesteps to get this percentage. Thus, the number of values included in these calculations is the number of selected event runs. Except for the average percentage which includes only detected events, these statistical calculations include all events. *Missed events* have 0% of timesteps alerting.
  - **Average Percent of Event Timesteps the EDS Alerts On for Detected Simulated Events**
  - **Minimum Percent of Event Timesteps the EDS Alerts On for All Simulated Events**
  - **First Quartile Percent of Event Timesteps the EDS Alerts On for All Simulated Events**
  - **Median Percent of Event Timesteps the EDS Alerts On for All Simulated Events**

- **Third Quartile Percent of Event Timesteps the EDS Alerts On for All Simulated Events**
- **Maximum Percent of Event Timesteps the EDS Alerts On for All Simulated Events**

### D.2.3 Analysis Settings

There are six user-specified analysis settings that impact the information produced in the ‘Analysis Data’ file. The user should verify the values before doing an ‘Analysis Data’ export. They are accessed via the ‘Edit’ Menu and are shown in Figure D-2.

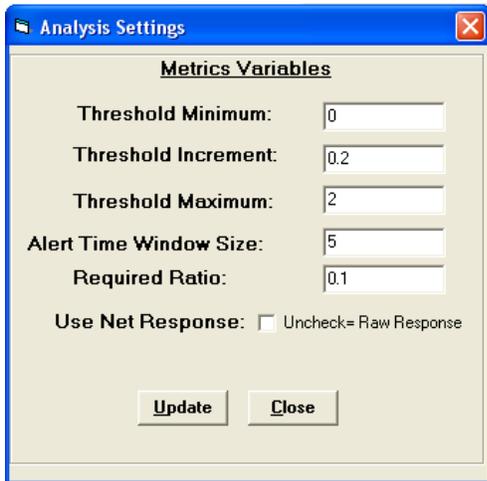


Figure D-2. Analysis Settings

Threshold minimum, increment and maximum do not impact calculations, but they determine which discrimination thresholds are reported in the ‘Analysis Data’ file. The first discrimination threshold for which calculations are included in this file is the threshold minimum, and then the threshold is increased by the threshold increment until the threshold maximum is reached. For example, if the minimum, increment and maximum thresholds are 0, 0.2 and 1 respectively, the file would include calculations for six discrimination thresholds: 0, 0.2, 0.4, 0.6, 0.8 and 1.

Alert time window size, required ratio and net response impact alert and detection calculations. They are discussed in Sections D.2.1 and D.2.2.

### D.3 Example Calculations

This section uses a simple example to illustrate how EDDIES calculates the key terms and central performance measures in the ‘Analysis Data’ and ‘Alerts’ export files. Section D.3.1 gives the data for example scenario to be used throughout this section and Sections D.3.2 – D.3.4 illustrate calculation of all metrics described in Section D.2. All performance metrics discussed in Section D.2 will be included in this section, though they are presented in a different order to improve clarity.

Section D.3 will use the “short hand” shown in Table D-1 to facilitate discussion.

**Table D-1. “Short hand” used in Section D.3**

Term	Short hand
Example Baseline Run Output, shown in Table D-2	Eg_Base
Example Simulated Event #1 Output, shown in Table D-3	Eg_Ev1
Example Simulated Event #2 Output, shown in Table D-4	Eg_Ev2
Timestep, referring to a particular timestep in the data (e.g., timestep 5 of Eg_Ev1)	Time
Timestep, referring to a count of timesteps (e.g., 3 timesteps were false positives)	TS
Discrimination Threshold	Threshold

**D.3.1 Example Scenario**

Tables D-2 through D-4 show the *event status* and EDS output for one *baseline dataset* and two *simulated event datasets* from the same example batch. It is assumed that the user has selected these three runs for export.

For the *simulated event runs*, only the output during the event period is shown, as these are the only timesteps considered by EDDIES during analysis. Note that for Eg\_Base, the *event status* was defined by the user in the imported dataset. For the event runs, the *event status* was assigned by EDDIES, with a “1” indicating that EDDIES modified the data for that timestep during event simulation.

**Table D-2. Example Baseline Run: “Eg\_Base”**

TIME_STEP	EVENT_STATUS	ALERT_LEVEL	ALERT_STATUS	TRIGGER_PARAMETERS
1	1	0.9	1	
2	0	0.8	1	
3	0	0.95	1	
4	0	0.9	1	
5	0	0.3	0	
6	-1	0.6	0	
7	-1	0.95	1	
8	-1	0.95	1	
9	0	0.3	0	
10	0	0.05	0	
11	0	0.05	0	
12	0	0.05	0	
13	0	0.1	0	
14	1	0.2	0	
15	1	0.25	0	
16	0	0.1	0	
17	0	0.05	0	
18	0	0.1	0	
19	0	0.35	0	
20	0	0.1	0	
21	0	0.15	0	
22	0	0.3	0	PH
23	1	0.5	0	PH
24	1	0.7	0	PH
25	1	0.85	1	PH
26	1	0.85	1	PH,TOC
27	1	0.95	1	PH,TOC
28	1	1.0	1	PH,TURB,TOC
29	0	1.0	1	PH,TOC

**Table D-3. Example Simulated Event 1: “Eg\_Ev1”**

TIME_STEP	EVENT_STATUS	ALERT_LEVEL	ALERT_STATUS	TRIGGER_PARAMETERS
10	1	0.85	1	CL2
11	1	0.9	1	CL2
12	1	0.85	1	CL2
13	1	0.95	1	CL2, COND
14	1	0.9	1	CL2, COND
15	1	0.8	1	CL2, COND

**Table D-4. Example Simulated Event 2: “Eg\_Ev2”**

TIME_STEP	EVENT_STATUS	ALERT_LEVEL	ALERT_STATUS	TRIGGER_PARAMETERS
17	1	0.15	0	
18	1	0.2	0	
19	1	0.45	0	
20	1	0.25	0	
21	1	0.25	0	

### **D.3.2 Calculation of Metrics Independent of Discrimination Threshold**

This section demonstrates how EDDIES calculates metrics discussed in Section D.2 that are not dependent on a *discrimination threshold* or *alert status*. Since they do not depend on a threshold, there is only one entry per file.

- **Number of Timesteps Identified as Normal by User in Imported data: 17.** There are 17 TSs in Eg\_Base with EVENT\_STATUS = 0.
- Statistical summary of alert level on normal timesteps: mathematical calculations based on the alert levels of these 17 timesteps, which are 0.8, 0.95, 0.9, 0.3, 0.3, 0.05, 0.05, 0.05, 0.1, 0.1, 0.05, 0.1, 0.35, 0.1, 0.15, 0.3, 1.
  - **Average Alert Level on Normal Timesteps: 0.33**
  - **Minimum Alert Level on Normal Timesteps: 0.05**
  - **First Quartile Alert Level on Normal Timesteps: 0.1**
  - **Median Alert Level on Normal Timesteps: 0.15**
  - **Third Quartile Alert Level on Normal Timesteps: 0.35**
  - **Maximum Alert Level on Normal Timesteps: 1.0**
  - **Standard Deviation of Alert Level on Normal Timesteps: 0.3**
- **Number of Abnormal Events: Identified by User in Imported data: 3.** There are three continuous sequences of event timesteps identified in Eg\_Base: a one-TS event at Time 1, a two-TS event from Times 14 - 15, and a six-TS event from Times 23 – 28.
- **Total Number of Timesteps in User-Identified Events: 9.** There are 9 TS with EVENT\_STATUS = 1 in Eg\_Base.
- Statistical summary of alert level on baseline event timesteps: mathematical calculations based on the alert levels of these 9 timesteps, which are 0.9, 0.2, 0.25, 0.5, 0.7, 0.85, 0.85, 0.95, 1.0.
  - **Average Alert Level for User-Identified Events: 0.69**
  - **Minimum Alert Level for User-Identified Events: 0.2**
  - **First Quartile Alert Level for User-Identified Events: 0.5**

- **Median Alert Level for User-Identified Events: 0.85**
- **Third Quartile Alert Level for User-Identified Events: 0.9**
- **Maximum Alert Level for User-Identified Events: 1.0**
- **Standard Deviation Alert Level for User-Identified Events: 0.3**
- **Number of Timesteps Identified as Bad Data by User in Imported data: 3.** There are 3 TSs in Eg\_Base with EVENT\_STATUS = -1.
- **Number of Simulated Events: 2.** This is always equal to the number of event runs selected, which in this case were Eg\_Ev1 and Eg\_Ev2.
- **Total Number of Timesteps in Simulated Events: 11.** There are 6 event TS in Eg\_Ev1 + 5 event TS in Eg\_Ev2.

Tables D-5 and D-6 show the net response for each timestep of Eg\_Ev1 and Eg\_Ev2. The *alert levels* are transferred from Tables D-2, D-3 and D-4. The net response is the difference between these values.

**Table D-5. Eg\_Ev1 Net Response Calculations**

TIME_STEP	ALERT LEVEL		Net Response
	Eg_Ev1	Eg_Base	
10	0.85	0.05	0.8
11	0.9	0.05	0.85
12	0.85	0.05	0.8
13	0.95	0.1	0.85
14	0.9	0.2	0.7
15	0.8	0.25	0.55

**Table D-6. Eg\_Ev2 Net Response Calculations**

TIME_STEP	ALERT LEVEL		Net Response
	Eg_Ev2	Eg_Base	
17	0.15	0.05	0.1
18	0.2	0.1	0.1
19	0.45	0.35	0.1
20	0.25	0.1	0.15
21	0.25	0.15	0.1

- Statistical summary of net response. The following metrics are mathematical calculations on the net responses from the 11 TSs within these two simulated events, shown in Tables D-5 and D-6.
  - **Average Net Response for Simulated Events: 0.46**
  - **Minimum Net Response for Simulated Events: 0.1**
  - **First Quartile Net Response for Simulated Events: 0.1**
  - **Median Net Response for Simulated Events: 0.55**
  - **Third Quartile Net Response for Simulated Events: 0.8**
  - **Maximum Net Response for Simulated Events: 0.85**

### **D.3.3 Alert Metrics**

There are several steps in determining alerts within EDDIES, as they depend on multiple variables and export characteristics selected by the user. Thus, this section breaks the discussion into two parts:

Section D.3.3.1 shows how EDDIES determines if a timestep is alerting and Section D.3.3.2 describes how this is then used to identify alerts.

### D.3.3.1 Alerting Timesteps

There are two primary methods for determining if a timestep is an alerting timestep. See Section 7 for details on how to select the desired method during export.

- Applying a discrimination threshold to the EDS-outputted alert level: if the timestep's *alert level*  $\geq$  *discrimination threshold*, the timestep is an *alerting timestep*.

This method can be selected for both 'Alerts' and 'Analysis Data' exports. For 'Alerts' exports, the user enters a single discrimination threshold to be used. For 'Analysis Data' exports, multiple thresholds are applied by EDDIES based on the user-defined range and threshold increment.

- Using the EDS-outputted alert status: if *Alert status* = 1, the timestep is an *alerting timestep*. This method is only an option for 'Alerts' exports.

These methods are illustrated in Tables D-7 – D-9, which use the example datasets shown in Tables D-2 – D-4. The first three columns of each table repeat the EDS output from these datasets. The next two columns show which timesteps would be considered *alerting timesteps* if the first method were used: determinations are shown for two discrimination threshold values. The final column shows the determinations if the user chose to analyze based on the alert status output.

**Table D-7. Alerting Timesteps for Eg\_Base**

<b>TIME_STEP</b>	<b>ALERT_LEVEL</b>	<b>ALERT_STAT_US</b>	<b>Alerting Timestep using Discrimination Threshold = 0.75</b>	<b>Alerting Timestep using Discrimination Threshold = 1.0</b>	<b>Alerting Timestep using ALERT_STATUS</b>
1	0.9	1	√		√
2	0.8	1	√		√
3	0.95	1	√		√
4	0.9	1	√		√
5	0.3	0			
6	0.6	0			
7	0.95	1	√		√
8	0.95	1	√		√
9	0.3	0			
10	0.05	0			
11	0.05	0			
12	0.05	0			
13	0.1	0			
14	0.2	0			
15	0.25	0			
16	0.1	0			
17	0.05	0			
18	0.1	0			
19	0.35	0			
20	0.1	0			
21	0.15	0			
22	0.3	0			
23	0.5	0			
24	0.7	0			
25	0.85	1	√		√
26	0.85	1	√		√
27	0.95	1	√		√
28	1.0	1	√	√	√
29	1.0	1	√	√	√

**Table D-8. Alerting Timesteps for Eg\_Ev1**

<b>TIME_STEP</b>	<b>ALERT_LEVEL</b>	<b>ALERT_STAT_US</b>	<b>Alerting Timestep using Discrimination Threshold = 0.75</b>	<b>Alerting Timestep using Discrimination Threshold = 1.0</b>	<b>Alerting Timestep using ALERT_STATUS</b>
10	0.85	1	√		√
11	0.9	1	√		√
12	0.85	1	√		√
13	0.95	1	√		√
14	0.9	1	√		√
15	0.8	1	√		√

**Table D-9. Alerting Timesteps for Eg\_Ev2**

TIME_STEP	ALERT_LEVEL	ALERT_STAT_US	Alerting Timestep using Discrimination Threshold = 0.75	Alerting Timestep using Discrimination Threshold = 1.0	Alerting Timestep using ALERT_STATUS
17	0.15	0			
18	0.2	0			
19	0.45	0			
20	0.25	0			
21	0.25	0			

There is one final option for determining alerting timesteps. For *simulated event datasets*, the user can choose to use the *net response* instead of the raw *alert level* for ‘Analysis Data’ export calculations. If this option is chosen, the following criteria will be used to determine alerting timesteps for *simulated event dataset(s)*: if  $Net\ response \geq discrimination\ threshold$ , the timestep is an *alerting timestep*. Even if this is selected, analysis of the *baseline dataset(s)* will always be characterized using the *alert level*. Thus, for each discrimination threshold in the ‘Analysis Data’ export file, the baseline metrics will be based on the *alert level* and the simulated event metrics will be determined based on the *net response*.

Tables D-10 and D-11 show which timesteps would be considered alerting if the *net response* were used for calculations, showing the same discrimination thresholds that were used in Tables D-7, D-8 and D-9. The *net response* values are taken from Tables D-5 and D-6.

**Table D-10. Eg\_Ev1 Alerting Timestep Calculations based on Net Response**

TIME_STEP	Net Response	Alerting Timestep using Discrimination Threshold = 0.75	Alerting Timestep using Discrimination Threshold = 1.0
10	0.8	√	
11	0.85	√	
12	0.8	√	
13	0.85	√	
14	0.7		
15	0.55		

**Table D-11. Eg\_Ev2 Alerting Timestep Calculations based on Net Response**

TIME_STEP	Net Response	Alerting Timestep using Discrimination Threshold = 0.75	Alerting Timestep using Discrimination Threshold = 1.0
17	0.1		
18	0.1		
19	0.1		
20	0.15		
21	0.1		

Note that with a discrimination threshold of 0.75, TSs 14 and 15 in Eg\_Ev1 were *alerting timesteps* when the raw *alert level* was used to determine alerting TSs (in Table D-8) but are not alerting when the *net response* is used (Table D-10).

The following example calculations are based on Tables D-7 through D-11.

- **Alerting Timesteps:**

- Using a *discrimination threshold* of 0.75 and not the *net response*: 11 in Eg\_Base + 6 in Eg\_Ev1, and 0 in Eg\_Ev2 = **17** alerting timesteps
- Using a *discrimination threshold* of 1 and not the *net response*: 2 in Eg\_Base + 0 in Eg\_Ev1, and 0 in Eg\_Ev2 = **2** alerting timesteps
- Using a *discrimination threshold* of 0.75 and the *net response*: 11 in Eg\_Base + 4 in Eg\_Ev1, and 0 in Eg\_Ev2 = **15** alerting timesteps
- Using a *discrimination threshold* of 1 and the *net response*: 2 in Eg\_Base + 0 in Eg\_Ev1, and 0 in Eg\_Ev2 = **2** alerting timesteps

\* For all subsequent analyses, calculations will be based on a *discrimination threshold* of 0.75 and will assume that the user has not chosen to use the *net response* in calculations.

*False positive* (FP) and *true negative* (TN) timesteps are calculated for normal TSs and thus only consider *baseline dataset(s)*. Table D-12 repeats the event status for each timestep in Eg\_Base, as well as the alerting status determined in Table D-7. The normal timesteps are classified as either FP or TN. There is a note as well for the alerting bad quality TSs.

**Table D-12. Normal Timestep Classification for Eg\_Base**

TIME_STEP	EVENT_STAT_US	Alerting Timestep using Discrimination Threshold = 0.75	Classification
1	1	√	-
2	0	√	FP
3	0	√	FP
4	0	√	FP
5	0		TN
6	-1		-
7	-1	√	Bad Quality
8	-1	√	Bad Quality
9	0		TN
10	0		TN
11	0		TN
12	0		TN
13	0		TN
14	1		-
15	1		-
16	0		TN
17	0		TN
18	0		TN
19	0		TN
20	0		TN
21	0		TN
22	0		TN
23	1		-
24	1		-
25	1	√	-
26	1	√	-
27	1	√	-
28	1	√	-
29	0	√	FP

- **Number of False Positive Timesteps - EDS incorrectly Alerting for normal timestep: 4**
- **Number of True Negative Timesteps - EDS correctly not Alerting for normal timestep: 13**
- **Percent of Normal Timesteps that are False Positives: 4 FP/17 normal timesteps = 23.53%**
- **Number of Timesteps with Bad Quality Data that the EDS Alerts on: 2**

#### *D.3.3.2 Determining Alerts*

Once *alerting timesteps* are identified, they can be translated into alerts and used to calculate the alerting metrics. The alert determination logic described in this section is used in both the ‘Alerts’ and ‘Analysis Data’ exports. All calculations below will be based on a discrimination threshold of 0.75 and no use of net response.

Alerts are determined using the user-defined alert time window (see Section 7.2 for instructions for setting these). Tables D-13 and D-14 show alerts for two alert time window sizes: 0 TSs and 2 TSs. Eg\_Ev2 is not included in this section as there were no alerting timesteps, and thus no alerts, in that dataset.

**Table D-13. Eg\_Base Alert Determination for a Discrimination Threshold of 0.75**

Timestep	Alerting Timestep using Discrimination Threshold = 0.75	EVENT_STAT_US	Alert Time Window Size (timesteps)	
			0	2
1	√	1	<b>Alert – Valid</b>	<b>Alert – Valid</b>
2	√	0	(Alert continued)	(Alert continued)
3	√	0	(Alert continued)	(Alert continued)
4	√	0	(Alert continued)	(Alert continued)
5		0		
6		-1		
7	√	-1	<b>Alert – Bad Quality</b>	(Alert continued)
8	√	-1	(Alert continued)	(Alert continued)
9		0		
10		0		
11		0		
12		0		
13		0		
14		1		
15		1		
16		0		
17		0		
18		0		
19		0		
20		0		
21		0		
22		0		
23		1		
24		1		
25	√	1	<b>Alert – Valid</b>	<b>Alert – Valid</b>
26	√	1	(Alert continued)	(Alert continued)
27	√	1	(Alert continued)	(Alert continued)
28	√	1	(Alert continued)	(Alert continued)
29	√	0	(Alert continued)	(Alert continued)

**Table D-14. Eg\_Ev1 Alert Determination for a Discrimination Threshold of 0.75**

Timestep	Alerting Timestep using Discrimination Threshold = 0.75	EVENT_STAT_US	Alert Time Window Size (timesteps)	
			0	2
10	√	1	<b>Alert – Valid</b>	<b>Alert – Valid</b>
11	√	1	(Alert continued)	(Alert continued)
12	√	1	(Alert continued)	(Alert continued)
13	√	1	(Alert continued)	(Alert continued)
14	√	1	(Alert continued)	(Alert continued)
15	√	1	(Alert continued)	(Alert continued)

- **Number of Invalid Alerts**
  - Window size = 0 TS: 0
  - Window size = 2 TS: 0
- **Invalid Alert Frequency (average number of timesteps between Invalid Alerts)**
  - Window size = 0 TS: NA (since there are no invalid alerts)
  - Window size = 2 TS: NA

- Statistical summary of invalid alert length
  - **Average Invalid Alert Length (timesteps):** NA (since there are no invalid alerts)
  - **Minimum Invalid Alert Length (timesteps):** NA
  - **First Quartile Invalid Alert Length (timesteps):** NA
  - **Median Invalid Alert Length (timesteps):** NA
  - **Third Quartile Invalid Alert Length (timesteps):** NA
  - **Maximum Invalid Alert Length (timesteps):** NA
- **Number of Alerts on Abnormal Data (Valid Alerts):** (only considers baseline data)
  - Window size = 0 TS: 2
  - Window size = 2 TS: 2
- Baseline data valid alert length (not included in ‘Analysis Data’ export file, but added to illustrate calculation of alert length)
  - Window size = 0 TS: 2 alerts with lengths 4 TS, 5 TS
  - Window size = 2 TS: 2 alerts with lengths 8 TS, 5 TS
- **Number of Alerts on Bad Quality Data**
  - Window size = 0 TS: 1
  - Window size = 2 TS: 0.

### D.3.4 Detection Metrics

All metrics related to event detection are dependent on the user-specified required ratio. Table D-15 illustrates the detection metrics for two required ratios: 0.001 and 0.75. Each of the three distinct event periods in Eg\_Base are considered separately.

**Table D-15. Determination of Events Detected for three Required Ratios**

Event	# Event Timesteps	# Timesteps Alerting	% Event Timesteps Alerting	Required Ratio			
				0.001		0.75	
				Event Detected?	# Detected Timesteps	Event Detected?	# Detected Timesteps
Eg_Base Event 1 (Time 1)	1	1	100%	Yes	1	Yes	1
Eg_Base Event 2 (Times 14-15)	2	0	0%	No	0	No	0
Eg_Base Event 3 (Times 23-28)	6	4	67.67%	Yes	4	No	0
Eg_Ev1	6	6	100%	Yes	6	Yes	6
Eg_Ev2	5	0	0%	No	0	No	0

- **Number of User-Identified Events Detected (True Positives)** (from Eg\_Base)
  - Required Ratio = 0.001: 2
  - Required Ratio = 0.75: 1
- **Number of User-Identified Events Not Detected (False Negatives)**
  - Required Ratio = 0.001: 1

- Required Ratio = 0.75: 2
- **Percent of User-Identified Events Detected**
  - Required Ratio = 0.001: 2 detected/3 events = 66.67%
  - Required Ratio = 0.75: 1 detected/3 events = 33.33%
- **Number of Detected Timesteps in User-Identified Events**
  - Required Ratio = 0.001: 1 + 0 + 4 = 5
  - Required Ratio = 0.75: 1 + 0 + 0 = 1
- **Number of True Positives - Simulated Events Detected** (from Eg\_Ev1 and Eg\_Ev2)
  - Required Ratio = 0.001: 1
  - Required Ratio = 0.75: 1
- **Number of False Negatives - Simulated Events Not Detected**
  - Required Ratio = 0.001: 1
  - Required Ratio = 0.75: 1
- **Percent of Simulated Events Detected**
  - Required Ratio = 0.001: 1 detected/2 events = **50%**
  - Required Ratio = 0.75: 1 detected/2 events = **50%**
- **Number of Detected Timesteps for Simulated Events**
  - Required Ratio = 0.001: 6 + 0 = 6
  - Required Ratio = 0.75: 6 + 0 = 6
- **Percent of Events Detected - User-Identified and Simulated Events** (all events)
  - Required Ratio = 0.001: 3 detected/5 events = 60%
  - Required Ratio = 0.75: 2 detected/5 events = 40%
- Statistical summary of percentage of event timesteps alerting for user-identified events (in Eg\_Base), which are 100%, 0% and 66.67%
  - **Average Percent of Event Timesteps the EDS Alerts On for Detected User-Identified Events: 33.56%**
  - **Minimum Percent of Event Timesteps the EDS Alerts On for All User-Identified Events: 0%**
  - **First Quartile Percent of Event Timesteps the EDS Alerts On for All User-Identified Events: 33.33%**
  - **Median Percent of Event Timesteps the EDS Alerts On for All User-Identified Events: 66.67%**
  - **Third Quartile Percent of Event Timesteps the EDS Alerts On for All User-Identified Events: 50.33%**
  - **Maximum Percent of Event Timesteps the EDS Alerts On for All User-Identified Events: 100%**

- Statistical summary of percentage of event timesteps alerting for simulated events (in Eg\_Ev1 and Eg\_Ev2), which are 100% and 0%.
  - **Average Percent of Event Timesteps the EDS Alerts On for Detected Simulated Events: 50%**
  - **Minimum Percent of Event Timesteps the EDS Alerts On for All Simulated Events: 0%**
  - **First Quartile Percent of Event Timesteps the EDS Alerts On for All Simulated Events: 25%**
  - **Median Percent of Event Timesteps the EDS Alerts On for All Simulated Events: 50%**
  - **Third Quartile Percent of Event Timesteps the EDS Alerts On for All Simulated Events: 75%**
  - **Maximum Percent of Event Timesteps the EDS Alerts On for All Simulated Events: 100%**

For ease of presentation, all subsequent calculations will assume a required ratio of 0.75.

Table D-16 shows the time to detect for each event. ‘ND’ indicates that the event was not detected.

**Table D-16. Determination of Time to Detect when Required Ratio = 0.75**

Event	Event Detected when Required Ratio = 0.75?	Time to Detect
Eg_Base Event 1 (Time 1)	Yes	0 (first TS is alerting)
Eg_Base Event 2 (Times 14-15)	No	ND
Eg_Base Event 3 (Times 23-28)	No	ND <sup>1</sup>
Eg_Ev1	Yes	0 (first TS is alerting)
Eg_Ev2	No	ND

<sup>1</sup> If Eg\_Base Event 3 had enough alerting timesteps to be detected with this required ratio, the time to detect would be 2, as there are 2 non-alerting timesteps before the first alerting timestep

- Statistical summary of time to detect for user-defined events (in Eg\_Base), which are 0, ND and ND.
  - **Average Time to Detect for Detected User-Identified Events (timesteps): 0 TS/1 detected event = 0 TS**
  - **Minimum Time to Detect for All User-Identified Events (timesteps): 0 TS**
  - **First Quartile Time to Detect for All User-Identified Events (timesteps):** this would theoretically be between 0 and ND, so ND
  - **Median Time to Detect for All User-Identified Events (timesteps): ND**
  - **Third Quartile Time to Detect for All User-Identified Events (timesteps): ND**
  - **Maximum Time to Detect for All User-Identified Events (timesteps): ND**
- Statistical summary of time to detect for simulated events (in Eg\_Ev1 and Eg\_Ev2), which are 0 and ND.

- **Average Time to Detect for Detected Simulated Events (timesteps):** 0 TS/1 detected event = **0 TS**
- **Minimum Time to Detect for All Simulated Events (timesteps):** **0 TS**
- **First Quartile Time to Detect for All Simulated Events (timesteps):** this would theoretically be between 0 and ND, so **ND**
- **Median Time to Detect for All Simulated Events (timesteps):** **ND**
- **Third Time to Detect for All Simulated Events (timesteps):** **ND**
- **Maximum Time to Detect for All Simulated Events (timesteps):** **ND**
- **Median Time to Detect for All Events - User-Identified and Simulated Events (timesteps):** median of 0, ND, ND, 0 and ND = **ND**
- **Average Time to Detect for Detected Events - User-Identified and Simulated Events (timesteps):** average of 0 and 0 = **0 TS**

The final metric that the user can choose to have calculated for detected events is trigger accuracy. This is calculated using *detected simulated events*. In the examples in this section, there is only one detected simulated event: Eg\_Ev1. Table D-17 shows the trigger parameters outputted by the EDS for the detected event timesteps (transferred from Table D-3).

**Table D-17. Trigger Parameters for Eg\_Ev1**

TIME_STEP	Alerting Timestep?	TRIGGER_PARAMETERS
10	√	CL2
11	√	CL2
12	√	CL2
13	√	CL2, COND
14	√	CL2, COND
15	√	CL2, COND

Below, the trigger accuracy for this event is calculated for two different contaminants.

- Trigger accuracy for the event if CL2, COND and PH were modified by EDDIES during event simulation: in this case, the EDS correctly identified two of the modified parameters (CL2 and COND), but never outputted the final parameter (pH) during the event. Thus, the trigger accuracy for this event would be  $2/3 = 66.67\%$ .
- Trigger accuracy for the event if only TOC was modified by EDDIES during event simulation: the EDS never outputted TOC during the event, so the accuracy for this event would be  $0/1 = 0\%$ .

Assuming the first example above in which CL2, COND and PH are modified, the average trigger accuracy shown in the 'Analysis Data' export is given below.

- **Average Trigger Accuracy:** Since there was only one detected simulated event, the average trigger accuracy would be this value of **66.67%**.

## Appendix E: Setpoint Analysis

Setpoint analysis is a feasible EDS option for most utilities as most SCADA and data management systems have the ability to produce alerts when parameter values surpass parameter setpoints (or thresholds) defined by the user. However, the setpoint values should be re-examined to ensure they are appropriate for detecting events.

EDDIES contains two tools to support utilities implementing setpoint analysis.

- The built-in Setpoint Algorithm allows the user to evaluate this analysis technique like any other EDS, creating batches for individual monitoring locations using utility data and running the defined datasets through this EDS.
- The Setpoint Threshold Analysis tool allows the user to compare performance for various threshold values for a single parameter. The user can create evaluation batches like with any other EDS evaluation.

This section describes each capability in more detail and includes instructions for setting up evaluations and interpreting results for each.

### E.1 Setpoint Algorithm

Like other EDSs, the Setpoint Algorithm produces alerts based on all data streams associated with a particular location. Its output is based on the HI and LO setpoint values defined for each parameter, as described in Section 4.

The Setpoint Algorithm output is described below.

- Alert level: 1 if *any* parameter surpasses its defined threshold value for the timestep and 0 otherwise.
- Alert status: the same as the alert level
- Trigger parameters: for alerting timesteps, parameter type(s) that are out of range are outputted

Table E-1 gives an example of the algorithm's analysis. The example assumes that the algorithm is monitoring three parameters at the given monitoring station: chlorine, conductivity and turbidity. The first four columns shows sample data for nine timesteps, and the last three columns show the resulting output that would be produced.

The user-defined setpoint values are at the top of the table, highlighted in peach. These setpoints require that chlorine be above 0.4, conductivity be between 300 and 500, and turbidity be less than 5. The chlorine HI setpoint and turbidity LO setpoint are blank/null and thus no alerts will be produced for high chlorine or low turbidity values.

**Table E-1. Example Data and Setpoint Algorithm Output**

SETPOINT VALUES	LO = 0.4, HI =	LO = 300, HI = 500	LO = , HI = 5			
TIME_STEP	A_CL2	A_COND	A_TURB	ALERT_LEVEL	ALERT_STATUS	TRIGGER PARAMETERS
1/1/2011 8:00	1.20	423	1.2	0	0	
1/1/2011 8:02	1.08	461	1.4	0	0	
1/1/2011 8:04	0.95	470	1.8	0	0	
1/1/2011 8:06	0.76	480	2.7	0	0	
1/1/2011 8:08	0.61	<i>500</i>	3.5	1	1	COND
1/1/2011 8:10	0.47	<i>530</i>	4.5	1	1	COND
1/1/2011 8:12	<i>0.37</i>	<i>535</i>	4.8	1	1	CL2, COND
1/1/2011 8:14	<i>0.39</i>	<i>525</i>	4.5	1	1	CL2, COND
1/1/2011 8:16	0.41	<i>514</i>	4.1	1	1	COND

Water quality values that surpass the defined setpoint values are shown in red and italicized. There are two chlorine values that are below the low setpoint of 0.4 and five timesteps that surpass the conductivity maximum of 500. No turbidity values fall outside the defined setpoint values during this period.

As noted above, the algorithm output is binary. As can be seen in this table, the alert level and alert status are 1 for timesteps in which any parameter surpassed a setpoint (8:08 – 8:16). The parameter type of all parameters outside of the acceptable range are also outputted. For the other timesteps, the alert level and alert status are 0 and no parameter types are exported.

Section 4.1.2 describes how setpoints are defined for each parameter. A few things to note:

- These HI and LO values are set per parameter ID, not per parameter type. For example, Setpoint LO for StationA\_CL2 could be 0.5 while Setpoint LO for StationB\_CL2 could be 0.2.
- HI and LO setpoints are not required for all parameters. Setpoint values should only be entered for parameters and values desired to generate alerts. The Setpoint Algorithm will not alert for a parameter if setpoints are not identified. For example, the utility chose not to define a LO setpoint in the example in Table E-1 because they were not concerned about low turbidity.
- \* A null setpoint is NOT the same as a setpoint of 0. If a setpoint HI is left null, for example, the parameter value can be infinitely high and no alarm would sound. However, if ‘0’ is entered in the SETPOINT HI field, the algorithm will alarm whenever the parameter value is greater than or equal to 0.

### ***E.1.1 Creating a Batch Using the Setpoint Algorithm***

The procedures for setting up and executing batches using the setpoint algorithm are the same as any other EDS. The following notes are important to review, however.

- The Setpoint Algorithm’s output depends entirely on the HI and LO setpoint values defined for each parameter. Thus, the user should verify the setpoint values for each parameter before creating or executing a batch using the setpoint algorithm. Section 4.1.2 describes how parameter setpoint values are viewed and edited.
- Like with other EDSs, all parameters associated with a monitoring location should be assigned to that location, as described in Section 4.1.4.

- When using the Setpoint Algorithm, the steps in Section 5 can be completely skipped: this EDS requires no installation or configuration.
- When creating a batch as described in Section 6.1, “Setpoint Algorithm” should be selected as the Configuration ID.

Once a batch has been executed with a given parameter, its setpoint values *cannot* be edited. This ensures that the results in the database are consistent with the current parameter settings. However, it makes comparison of multiple setpoint values somewhat cumbersome. Three options for evaluating more than one setpoint value for a given parameter are described below.

The easiest way to compare performance with more than one setpoint value is to use the setpoint sensitivity analysis tool described in Section E.2. While this only allows analysis of one parameter at a time, it is robust and easily implemented.

If the user wishes to consider all parameters together, one option to overcome this restriction is to create separate database instances for each set of setpoint values. Appendix F describes the scripts for exporting database results (export.bat), clearing the database (deploy.bat), and importing a previously generated database instance (import.bat).

Use of database scripts, however, requires that the user complete all setup steps again, including defining parameters and locations and importing data. If the user knows what analyses and information exports they will want for each batch, EDS results only can be cleared between batches. In this solution, the user creates and executes all batches containing the first set of setpoint values and then exports ALL data, results and analysis files that could potentially be used based on these setpoint values. Once all information is captured, results for all batch(es) for which other setpoint values are desired are deleted by right-clicking the batch name(s) on the Launch Manager and selecting “Delete Results”. The setpoint values can then be modified and the batch(es) can be re-executed and the new results can be exported.

The other option for considering all parameters together is more cumbersome, but the results remain in the database and can thus be queried at any time. In this solution, separate parameters are created for each desired setpoint value. For example, if a user wanted to evaluate 0.2 and 0.5 as Station A’s chlorine LO setpoint, they could define the parameters A\_CL2\_LO\_0.2 and A\_CL2\_LO\_0.5 and assign the desired setpoint values for each. The user then must import utility data for each separately (presumably the same data but with different parameter IDs as the column header). Separate locations would also need to be created (perhaps Location\_A\_1 and Location\_A\_2) using the different sets of parameter.

### **E.1.2 Analyzing Results from the Setpoint Algorithm**

This section lists each EDDIES export type and describes its relevance related to analysis of setpoint algorithm results. Section 7 describes how to export these files and Appendix C gives the format and data included in each.

#### **E.1.2.1 Alert Export**

The ‘Alerts’ export can be useful for analyzing setpoint algorithm results. As with any EDS, this file contains all alerts generated using the user-defined setpoint values and indicates if each alert is a true positive, false positive or occurring during a period of bad data quality.

When doing an ‘Alerts’ export with a batch using the Setpoint Algorithm, the steps in Section 7.2.1 should be followed, and the user should select ‘No’ when asked if they want to enter a threshold to determine alerts.

### ***E.1.2.2 Results and Data Export***

Section 7.2.2 describes this file type. Creating this export file and interpreting its contents is exactly the same for the setpoint algorithm as any other EDS.

### ***E.1.2.3 Analysis Data Export***

While all metrics included in the ‘Analysis Data’ export can be calculated for batches using the setpoint algorithm, the binary nature of this EDS’s output means that 0 and 1 are the only threshold values with meaning. If the user wishes to have these metrics calculated, the following analysis variable values should be used (Section 7.2.3 describes how to do this). Note that this will create an export file with only two columns.

- Discrimination Threshold Increment = 1
- Discrimination Threshold Minimum = 0
- Discrimination Threshold Maximum = 1

### ***E.1.2.4 Setpoint Sensitivity Analysis Export***

The ‘Setpoint Sensitivity Analysis’ export is not relevant for batches using the setpoint algorithm.

## **E.2 Setpoint Sensitivity Analysis**

The Setpoint Sensitivity Analysis is quite different from other EDSs in that it considers only one parameter at a time. It is intended to support selection of setpoint values for event detection, allowing the user to see the valid and invalid alerts that would be produced at various setpoint settings.

The procedures for defining parameters and batches are different than with other EDSs, and interpretation of the export files is unique as well. This section describes how to use the setpoint sensitivity analysis capability and interpret the results.

### ***E.2.1 Using the Setpoint Sensitivity Analysis Tool***

The majority of steps required when setting up and executing a batch using the setpoint sensitivity analysis tool are the same as when using any other EDS. However, there are some differences for use of this tool, and these are described below.

- Unlike the setpoint algorithm described in Section E.1, the setpoint sensitivity analysis tool does *not* use the parameter setpoint values defined by the user and thus setpoint values do not need to be defined by the user.
- \* Because it is designed to evaluate a single parameter, each location must contain only one parameter. Thus, separate locations must be created for each individual parameter to be analyzed, as described in Section 4.1.3 and 4.1.4. For ease, it is suggested that these locations be given the name of the parameter (for example, name the location A\_CL2 if this is the parameter used).
- When using this tool, the steps in Section 5 can be completely skipped: this tool requires no installation or configuration.
- During batch creation, “Setpoint Sensitivity Analysis” should be selected as the Configuration ID.

Also, when creating a batch using the setpoint sensitivity analysis tool as described in Section 6.1, only contaminants impacting the parameter type being analyzed should be selected. Otherwise, the event

datasets would be exactly the same as the baseline run and thus no additional information would be received. For example, if the parameter A\_PH is being evaluated, only contaminants impacting PH should be selected. Section B.3.1 describes how to view which parameter types are impacted by each contaminant.

## ***E.2.2 Analyzing Results from the Setpoint Sensitivity Analysis Tool***

This section lists the export types within EDDIES and details the information that is provided and related to a setpoint sensitivity analysis for each type.

### ***E.2.2.1 Alert Export***

The ‘Alerts’ export was not specifically designed to evaluate results for the Setpoint Sensitivity Analysis tool. However, this export can be used to view alerts generated based on a specified HI setpoint value (essentially simulating the setpoint algorithm based on a specified HI setpoint). The ‘Alerts’ export will not generate alerts based on LO setpoint values.

Section 7.2.1 details the process for performing an ‘Alerts’ export. When implementing this export, the user must select ‘Yes’ when asked if they want to enter a threshold to determine alerts. They should then enter the desired HI setpoint value for the parameter. For example, if the batch is based on parameter A\_PH and the user enters a value of 9.0 for the threshold, the resulting Alerts export file would contain all alerts that would be generated if a HI setpoint value of 9.0 had been specified for this parameter.

### ***E.2.2.2 Results and Data Export***

This export type is not relevant for batches using the setpoint sensitivity analysis tool. If the user wishes to view the batch’s datasets, time-series files that contain that data can be exported via the ‘Launch Manager’ tab as described in Section 6.2.2.1.

### ***E.2.2.3 Analysis Data Export***

This export type is not relevant for batches using the setpoint sensitivity analysis tool. Instead, the ‘Setpoint Sensitivity Analysis’ export described in Section E.2.2.4 should be used, which contains the same metrics as the Analysis Data Export but is specifically designed to analyze and evaluate results generated by the Setpoint Sensitivity Analysis.

### ***E.2.2.4 Setpoint Sensitivity Analysis Export***

The Setpoint Sensitivity Analysis export was designed especially for the setpoint sensitivity analysis tool. The sections below describe the contents of these files, as well as the procedure for export.

#### ***E.2.2.4.1 File Contents***

The Setpoint Sensitivity Analysis Export file contains the same metrics as the Analysis Data export file, but instead of each column representing an EDS’s alerting threshold, each corresponds to performance that would be observed if the given setpoint value were used. Each analysis type is described below.

Tables E.2 and E.3 show sample ‘Setpoint Sensitivity Analysis’ export files. Note that every term in these files is also in the Analysis Results Export file, with two exceptions. “Setpoint Value” has replaced the term “Discrimination Threshold,” and the “Setpoint Type” row is added, which notes which selection the user made during export.

Appendix D describes these terms and metrics in detail and gives example calculations. The calculations are identical except that the criteria for determining alerting timesteps. For the Setpoint Sensitivity Analysis Export file:

- If the user chooses to evaluate high setpoints, a timestep is considered an alerting timestep if the parameter value for the timestep  $\geq$  the given setpoint value. Thus, a given timestep will be alerting for all setpoint values less than or equal to the parameter value. For example, if the TOC value for a given timestep is 0.8 ppm, this timestep would be considered alerting for all setpoint values less than or equal to 0.8.
- If the user chooses to evaluate low setpoints, a timestep is considered an alerting timestep if the parameter value for the timestep  $\leq$  the given setpoint value. Thus, a given timestep will be alerting for all setpoint values greater than or equal to the parameter value. For example, if the chlorine value for a given timestep is 0.8 mg/L, this timestep would be considered alerting if the setpoint value is greater than or equal to 0.8 (e.g., if the LO setpoint were 1.0, the value of 0.8 would trigger an alert).

An example interpretation of each file type is given below.

### **High Setpoint Analysis**

Table E-2 gives an example Setpoint Sensitivity Analysis Export file using a high setpoint. As noted above, a timestep is considered alerting if the parameter value is greater than or equal to the setpoint value.

**Table E-2. Sample High ‘Setpoint Sensitivity Analysis’ Export File**

Number of Simulated Event Runs in Analysis	10				
Number of Batches in Analysis	1				
Polling Interval	5				
Setpoint Type	High to detect increase in parameter value				
---ANALYSIS PARAMETER SETTINGS---					
Setpoint Value Increment	0.5				
Setpoint Value Minimum	0				
Setpoint Value Maximum	2				
Alert Time Window	5				
Use Net Response for Contamination Event Analysis	no				
Required Ratio for Detection	0.1				
---SUMMARY OF ANALYSIS DATA---					
Number of Timesteps Identified as Normal by User in Uploaded Data	8536				
Number of Timesteps Identified as Bad Data by User in Uploaded Data	55				
Number of Abnormal Events Identified by User in Uploaded Data	1				
Number of Simulated Events	10				
SETPOINT VALUE	0	0.5	1	1.5	2
***** OVERALL EDS PERFORMANCE *****					
Percent of Normal Timesteps that are False Positives	100%	99.99%	8.18%	0.02%	0.01%
Number of Invalid Alerts	1	2	4	1	1
Average Invalid Alert Length (timesteps)	8640	2431	63.82	2	1
Percent of Events Detected - User-Identified and Simulated Events	100%	100%	81.82%	27.27%	0%
Percent of User-Identified Events Detected	100%	100%	100%	0%	0%

Number of Simulated Event Runs in Analysis	10				
Percent of Simulated Events Detected	100%	100%	80%	30%	0%
Median Time to Detect for All Events - User-Identified and Simulated Events (timesteps)	0	0	20	ND	ND

Consider the column in Table E-2 showing performance when the setpoint value is 1.5. In this case, only 0.02% of timesteps are false positives: 0.02% of normal timesteps in the uploaded data are greater than or equal to 1.5. However, only 27% of events are detected: in only 27% of the events were parameter values greater than or equal to 1.5 for a sufficient time (10% of the event, per the required ratio shown in this file).

If the HI setpoint were lowered to 0.5 in this example (in which case the 0.5 column would be considered), the parameter value in almost all (99.99%) normal timesteps are alerting (as they're greater than or equal to 0.5). With this setpoint, all events would be detected.

Based on this export file, the user would likely choose a HI setpoint value of 1.0 for this parameter – as the majority of events (82%) were detected with this value, while only four invalid alerts were generated. Alternately, the user could do another Setpoint Sensitivity Analysis export with a smaller increment to further refine the value – perhaps with a setpoint value minimum of 0.5, maximum of 1.5 and increment of 0.1.

### **Low Setpoint Analysis**

Table E-3 shows a sample 'Setpoint Sensitivity Analysis' export file for analysis of LO setpoint values. Again, for the LO setpoint analysis, a timestep is alerting if the parameter value is lower than the given setpoint value. This example is unrelated to that in Table E-2.

**Table E-3. Sample Low 'Setpoint Sensitivity Analysis' Export File**

Number of Simulated Event Runs in Analysis	1				
Number of Batches in Analysis	1				
Polling Interval	2				
Setpoint Type	Low to detect decrease in parameter value				
---ANALYSIS PARAMETER SETTINGS---					
Setpoint Value Increment	1				
Setpoint Value Minimum	0				
Setpoint Value Maximum	4				
Alert Time Window	5				
Use Net Response for Contamination Event Analysis	no				
Required Ratio for Detection	0.1				
---SUMMARY OF ANALYSIS DATA---					
Number of Timesteps Identified as Normal by User in Uploaded Data	8536				
Number of Timesteps Identified as Bad Data by User in Uploaded Data	55				
Number of Abnormal Events Identified by User in Uploaded Data	1				
Number of Simulated Events	1				
SETPOINT VALUE					
	0	1	2	3	4
***** OVERALL EDS PERFORMANCE *****					
Percent of Normal Timesteps that are False Positives	0%	0.01%	95.12%	99.98%	99.99%
Number of Invalid Alerts	0	1	66	1	1
Average Invalid Alert Length (timesteps)	NA	1	896	8640	8640

Number of Simulated Event Runs in Analysis	1				
Percent of Events Detected - User-Identified and Simulated Events	0%	0%	100%	100%	100%
Percent of User-Identified Events Detected	0%	0%	100%	100%	100%
Percent of Simulated Events Detected	0%	0%	100%	100%	100%
Median Time to Detect for All Events - User-Identified and Simulated Events (timesteps)	ND	ND	4	0	0

With a LO setpoint export, the number of alerting timesteps increases as the setpoint value increases. All alerting timesteps at the lower setpoint would also be alerting at the higher ones (e.g., all water quality values less than one are also less than two).

In this case, it seems that most water quality values are between one and two, as very few timesteps are alerting with a setpoint value of one, but almost all are alerting with a setpoint value of two. As there are no setpoint values for which a reasonable number of invalid alerts are produced while most events are detected, it would definitely be recommended that the user do another export in this case with a smaller setpoint value increment, considering potential setpoint values between one and two.

#### E.2.2.4.2 Procedure for Exporting

The steps below should be followed to produce a ‘Setpoint Sensitivity’ file. Note that generation of these files can take a long time.

1. From the ‘Edit’ menu, select the ‘Analysis Variables’ button and verify the variables. These are described in Section D.2.3. The following suggestions are made for this export type.
  - Threshold Minimum and Threshold Maximum: A wider range than is normally observed should be selected, as simulated events will produce water quality values outside normal values. For example, if a utility typically has pH values between 8 and 9, they may choose 7 and 10 as their threshold minimum and maximum values.
  - Threshold Increment: For an initial export, it is recommended that an increment is selected that results in approximately 10 reported values. For example, if the minimum and maximum values were 7 and 10 (a spread of 3), an increment around 0.3 would be appropriate. More focused export files can be produced once a good range is identified.
  - Use Net Response: The box should *not* be checked.
2. Select the run(s) to be exported from the ‘Export Manager’ tab (as described in Section 7.1). For Setpoint Sensitivity Analysis exports, all selected runs should use the same location, and thus the same parameter.
3. Select the ‘Export Setpoint Sensitivity’ button on this tab.
4. On the ‘Setpoint Sensitivity Detection Threshold Setting’ window that pops up, specify if the export should analyze HI or LO parameter setpoint values.
  - Click ‘Yes’ to simulate a HI setpoint, in which an alert is produced if a water quality value is greater than or equal to the setpoint value.
  - Click ‘No’ to simulate a LO setpoint, in which an alert is produced if a water quality value falls below the setpoint value.

The progress of the export is shown across the bottom of the ‘Export Manager’ tab. A window will pop up when the export is complete.

## Appendix F: Database Scripts

This appendix details the four database scripts included in the EDDIES installation materials: the deploy database script, the deploy EDS database script, the export database script and the import database script.

The deploy and deploy EDS scripts are run during initial database setup, as described in Sections 3.2 and 5.1. The remaining scripts can be used to create a new database (the deploy script), save the information in the database (the export script), and reload a previously exported database (the import script).

These scripts are important if a user wishes to evaluate several EDSs, use large amounts of data, or run a lot of batches - particularly if the user is using the free version of Oracle (described in Section 3.1) which has a database size constraint. Using these scripts, the user can essentially keep database instances as separate “files” and only load the data currently in use.

### F.1 Deploy Database Script

The deploy database script (deploy.bat) creates a new EDDIES database, executing a series of Oracle scripts that creates the EDDIES tables, procedures, users and user permissions. Comparing the database to a word processing application, this script is essentially a “create new document” function.

The deploy script is executed during initial installation of EDDIES (as described in Section 3.2) and during disaster recovery to restore the system. It is also executed to create a new database and before importing a different database. Instructions for executing this script are given in Section 3.2.2. Details on the EDDIES tables and users it creates are available in the supplemental Database Design document.

\* Note that the deploy script erases any data previously in the EDDIES database. If the user has data in the EDDIES database, it is suggested that they run the export script first, as described in Section F.3. This script does not impact data in the Oracle database from other applications.

### F.2 Deploy EDS Database Script

The deploy EDS database script (deploy\_eds.bat) creates a new EDS account within EDDIES. Each EDDIES-compatible EDS (described in Section 5.1) must have a user account in the EDDIES database in order to obtain data from and post data to the database.

Section 5.2.2 describes how to execute the deploy EDS script. If the user wishes to change an EDS’s account information (change the password, for example), the deploy EDS script can be re-run, entering the updated information.

### F.3 Export Database Script

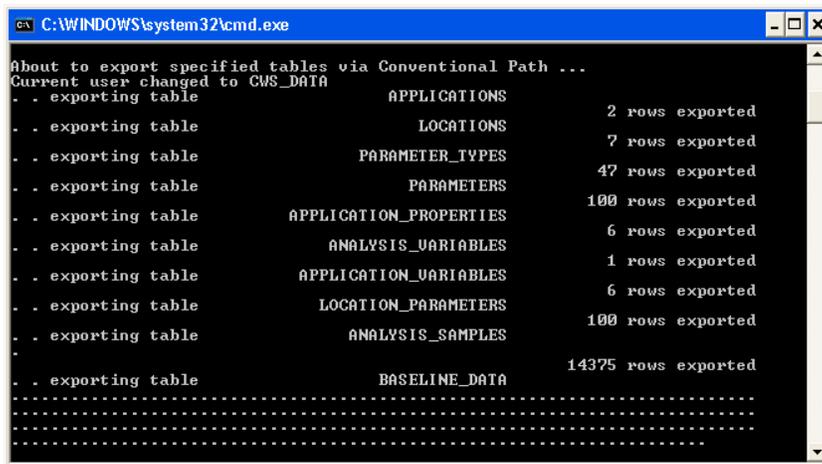
The export database script (export.bat) exports all data and tables in the EDDIES database into a dump (DMP) file. This file can be imported later using the import database script. Comparing the database to a word processing application, this script is essentially a “save document” function.

It is suggested that this script be run periodically to backup the database so the data can be restored if it becomes corrupted. Running the export script does not impact the database in any way.

The following steps detail procedures for executing the export database script.

1. Execute the export scripts using one of the following methods.
  - Navigate to the 'Export' folder found in the 'Script and DMP File' folder within in the EDDIES program folder. The default location is C:\Program Files\EDDIES 4.4\Script and DMP File. Execute the export database script by double clicking the 'export.bat' batch file.
  - From the Windows Start menu, go to All Programs | EDDIES 4.4 | Create Data Dump.
2. When prompted in the DOS window that pops up, enter the SYSTEM password chosen during Oracle installation (discussed in Section 3.1).

The DOS window shows the progress of the export, as shown in Figure F-1, and automatically closes when the export is complete. When the export is complete, this text is saved in the 'export.log' document in the 'Export' folder.



**Figure F-1. Export Progress**

3. Locate the 'eddiess.dmp' file in the 'Export' folder. Rename this file (for example, naming it 'eddiess\_July2013.dmp' or 'Setpoints\_1.dmp') and move it to a different folder if desired.

This file contains all data and tables from the export and can be imported at any time using the import database script. However, if it is not renamed or moved, it will be overwritten the next time the export database script is executed.

#### **F.4 Import Database Script**

The import database script (import.bat) imports a database DMP file previously created using the export database script. Comparing the database to a word processing application, this script is essentially an "open document" function.

\* The import database script *does not* delete previously added data. Therefore, the deploy database script should be run prior to the import database script to create a clean database for import.

Instructions for executing the import batch file are given in Section 3.2.2, under step #2.

## Appendix G: EDDIES Keyboard Shortcuts

This appendix describes the keyboard shortcuts that can be used for mouse-less navigation through EDDIES. Note that the keyboard shortcuts using the Alt key described in this section are not valid with the use of 'FilterKeys' in Windows Accessibility options, though the Tab key can be used instead to navigate through EDDIES.

### G.1 Main Menu Bar

Pressing and releasing the Alt key activates the keystrokes for the Main Menu bar, which contains the File, Add, Edit, View and Help Menus. Figure G-1 shows how the shortcut letter is underlined for each menu item once the Alt key has been pressed (the underlining does not show up and the links are not active otherwise).

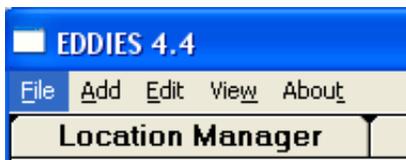


Figure G-1. Main Menu Bar with Shortcuts Active

The keystrokes for each menu item is listed below. Again, the Alt key must be pressed *and released* before pressing the letter.

- File Menu: Alt and then F
- Add Menu: Alt and then A
- Edit Menu: Alt and then E
- View Menu: Alt and then W
- About Menu: Alt and then T

Alternately, the Alt key can be pressed and released to select the main menu, and then the left and right arrow keys can be used to move between menu items.

### G.2 Navigation within an EDDIES Tab Screen

The field that is currently active is indicated by a box around the field name or a cursor in the field. Similar to Windows, the Tab key moves to the next hot item on a page.

- Advancing to the next item on an EDDIES page: Tab key
- Moving to the previous item: Shift + Tab

If the Tab key is pressed while the last item on an EDDIES page is highlighted, it will advance to the first item on the next EDDIES page. Likewise, pressing Shift + Tab when on the first item on an EDDIES page moves to the last item of the previous page.

Alternately, key strokes can be used to navigate directly to the desired item on the current EDDIES page. To advance directly to a dropdown list or button on a page, press the Alt + the letter underlined in the text

of the item's description. For example, on the Launch Manager page, press Alt + E to execute a batch that is selected, or on the Location Manager page, press Alt + D to add a parameter.

Below are descriptions for how to use the various field types with mouse-less interaction.

### **Check boxes**

Several tabs have boxes to check/uncheck. Navigate to these fields using tabbing or a hotkey, as described above. Once navigated to the desired entry, use the spacebar to check/uncheck the box.

### **Drop-down lists**

In several cases, drop-down lists are used from which users select from previously-defined items. After navigating to a drop-down list using a hotkey or via tabbing, the up and down arrow keys are used to move through the entries in the list.

### **Tree views**

Tree views are used on the Launch Manager and Export Manager tabs, shown in Figures 6-4 and 7-1. Navigate to the desired box using tabbing or hotkeys. Once inside, use the up and down arrows to move from one item to the next. For items with a plus sign in front of it, hit Enter to expand the item. Hitting Enter again collapses it back down.

## Appendix H: Common Issues

This appendix addresses the most common problems that users have experienced when setting up and/or running EDDIES.

- I get a warning message that “The destination file is in-use” during EDDIES installation.
  - ‘Ignore’ these error(s). It will not impact the installation of EDDIES.
- I cannot connect to the Oracle database, though the log-in information I entered is correct.
  - Stop the database and then start it back up. The stop and start options can be found by going to Start | All Programs | Oracle Database.
- I get the error “Listener does not know of service requested” when I open EDDIES.
  - Start the Oracle database. The start options can be found by going to Start | All Programs | Oracle Database.
- I’m getting an error from EDDIES saying “database connection lost”.
  - This is typically seen when the computer where EDDIES is being run has a wireless network/internet connection. Oracle fails when the internet connection goes on and off (this may be tied to the fact that the Database Home Page opens in the Internet Explorer). This problem should be resolved if the user disables the network connection.
- When I try to execute any action, I get a permissions error.
  - EDDIES must have administrator rights. In Windows 7, this can be done by right-clicking over the EDDIES program name from the Start Menu and then selecting “Run as Administrator”.
- Whenever I try to upload data or run a batch, I get an error and the process aborts.
  - You have likely reached the limit of the database storage. Use the Oracle interface to check the data storage status (see the Oracle documentation for instructions on how to access this information). If the limit has been reached, data will need to be deleted. One option is to export the current database and re-run the deploy.bat script. See Appendix F for more details on exporting and deploying the EDDIES database.
- I’m getting an error from EDDIES saying “Cannot start more than one transaction on this session”.
  - Close and restart EDDIES.
- I’m getting a “Path/File access error” when I try to import data.
  - Update permissions in the EDDIES 4.4 folder using the following steps:
    1. Navigate to C:\Program Files\EDDIES 4.4 (or whatever directory EDDIES 4.4 was installed in).
    2. Right click in empty space within the folder and select ‘Properties’.
    3. Go to the ‘Security’ tab in the window that appears.

4. If using Windows XP, click the 'Users' group in the box at the top of the window, ensure the 'Write' box is checked at the bottom of the window to Allow that functionality and click 'OK.'
  5. If using Windows 7, click the 'Edit' button and then follow the instructions given in step #4 above.
- The information I entered is not recognized. For example, I installed the EDS correctly but EDDIES is saying that it cannot find it.
    - The Oracle database is case-sensitive. If capital letters are used in one location and lower-case in another, EDDIES will assume these are different entries.
  - I am having problems when I create a parameter, parameter type, location, etc. that starts with a number.
    - It is suggested that all parameters, parameter types, locations, etc. start with a letter.
  - I've made a change (to a configuration, location...) but the change isn't showing up on the EDDIES User Interface.
    - Close and restart EDDIES.
  - There is an error regarding the number of columns when I try to import a file.
    - Open it in Notepad and ensure there are no extra commas or spaces at the end of the rows. If there are, go into Excel and delete the extra columns.
  - I'm getting an error regarding the date format of the file I'm trying to import.
    - Open it in Notepad and ensure the date is in the correct format.
  - When I try to import a file, I get an error that EDDIES cannot find 'loader.ctf'
    - If the file to be imported is on an external drive (e.g., a thumb drive), move it to your computer's hard drive and re-import.
  - When I try to do an analysis or setpoint sensitivity analysis export, I get the error "Threshold maximum is not a multiple of the threshold increment" even though the increment is correct.
    - Click OK and continue. The export file will be generated correctly.
  - The file I exported looks incorrect. The data seems mixed up and in the wrong format.
    - Only one file export can be done at a time. Re-export the file.
  - I have deleted results for a previously executed batch, but I am getting an error when I try to re-run the batch/re-import results.
    - Close and restart EDDIES.