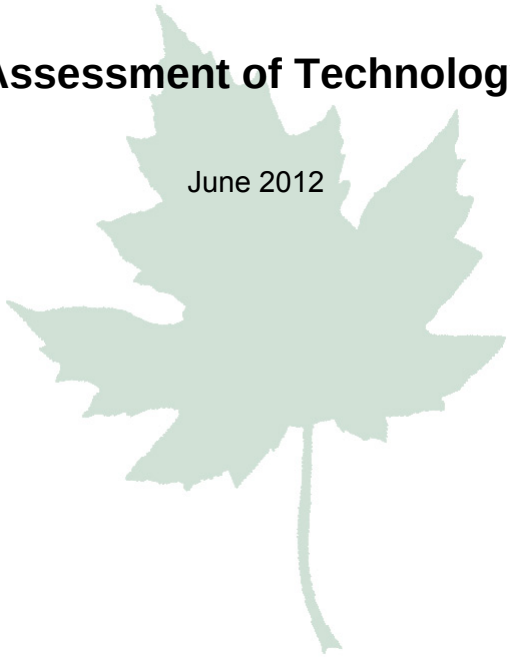
Environmental Technology Verification

General Verification Protocol (GVP)

Review of Application &

Assessment of Technology

June 2012



Environment Canada Environnement Canada

TABLE OF CONTENTS

TABLE OF CONTENTS 1
LIST OF FIGURES 3
LIST OF TABLES 3
LIST OF APPENDICES 4
EXECUTIVE SUMMARY 7
PART ONE –VERIFICATION BACKGROUND 9
1.0 Purpose and Structure of Verification 9
1.1 Definition of Verification 9
1.2 Canadian ETV Program Program 9
1.2.1 Background 9
1.2.2 ETV Program 9
1.3 Definitions 10
1.3.1 Environmental Technology 10
1.3.2 Equipment – based Service 10
1.3.3 ETV Fact Sheets 11
1.3.4 Verification Organizations (VOs) 11
1.3.5 Verification Certificate 11
1.3.6 Verification Report 11
1.5 Eligibility of Applicants for the ETV Program 14
1.5 Eligibility of Applicants for the ETV Program 15
1.6 Canadian ETV Program Program 15
PART TWO – VERIFICATION PROCEDURE 16
2.0 Review of Application Documents 16
2.1 Application Documents 16
2.2 Application Identification 16
2.3 Application Review Criteria 16
2.4 Qualifications and Independence of Test Agency and Analytical Laboratory 18
2.4.1 Qualifications and Independence of Test Agent 18
2.4.2 Qualifications and Independence of Analytical Laboratory 19
2.4.3 Test Programs that do not involve an Analytical Laboratory 19
2.5 Technology Type: Specific Verification Protocol & Supplemental Documents 19
2.6 Review of Conformance to Performance Benchmarking 19
3.0 Review of Technology 20
3.1 Scientific and Technical Foundation 20
3.2 Operation and Maintenance 21
3.3 Technology Review Criteria 21

4.0 Review of Test Plan, Test Execution and Data 26
4.1 Review of Test Plan and Execution of Test Plan 26
4.2 Review of Original Data 26
4.2.1 Data Set Identification 26
4.3 Review of Documentation from Other Institutions 27
4.4 Data Validity Checklist 30
5.0 STATISTICAL ANALYSIS 33
5.1 Formation of a Hypothesis Regarding Performance 33
5.1.1 The Null Hypothesis 33
5.1.2 Data Assessment 34
5.2 Degree of Confidence and Confidence Intervals 34
5.3 Statistical Analysis of the Performance Claim(s) 35
5.3.1 Performance Claim 35
5.3.2 Performance Claim(s) Verification 35
5.4 Data Analysis Checklist 35
5.5 Data Interpretation Checklist 36
5.6 Summary of Acceptable Datasets for Verification 38
5.7 Revision of Performance Claims 38
6.0 Alternate Means of Establishing Verified Performance 39
6.1 Vendors with Certification under other Jurisdictions 39
6.2 Technology That Has Been Recognized For Conformance To Science-Based Regulations 39
6.3 Participants in a Performance Benchmarking Program 40
6.4 Proof-of-Principle Programs for Pre-Commercial Technology 40
6.5 Technology With Test Data That Is Not Categorized Under the Canadian ETV Program SAW (Statistical Analysis Worksheet) 41
6.6 Harmonization With Programs Of Verification In Other Jurisdictions 41
6.7 Notes on related verifications [OPTIONAL] 41
7.0 The Verification Report 42
7.1 Establishment of the Audit Trail 42
7.1.1 Summary of key supporting documents 42
7.2 Example Report Format for Environmental Technology Verification Reporting 42

LIST OF FIGURES

Figure 1: The Verification Process 12
Figure 2: 14

LIST OF TABLES

Table 1: Application Review Checklist – Mandatory Information 17
Table 2: Technology Review Criteria Checklist 21
Table 3: Verification Study Design Assessment Criteria Checklist 28
Table 4a: Data Validity Checklist 31
Table 4b: Remote Sensing Data 33
Table 5: Data Analysis Checklist 36
Table 6: Data Interpretation Checklist 36
Table 7: Summary of Acceptable Data Sets for Verification 38
Table 8 Key documents 42

LIST OF APPENDICES

Appendix A: Statistical Analysis Worksheets
Appendix B: Selected Statistical Tables

Environmental Technology Verification
(ETV)
General Verification Protocol
June, 2012

Verification Protocol Authorization on behalf of the Canadian ETV Program

I. **Under License From**:

Environment Canada

II. **Issue History:**

First Issue: May, 1998
Revised: March, 2000
Revised: February, 2007
This Document: June, 2012, Replaces all previous issues
The ETV General Verification Protocol is Subject to Revision without Notice;

III. **Contact Information for Users of the ETV General Verification Protocol:**

Environment Canada,
Science and Technology Programs
Science and Technologies Strategies Directorate
Science and Technology Branch
200 Sacre-Coeur Blvd.
Gatineau (Hull), QC Canada K1A 0H3
Phone: 819 934 6419
jeffrey.guthrie@ec.gc.ca

EXECUTIVE SUMMARY

This protocol is a working document that provides step by step procedures for conducting a verification of performance for an environmental technology or process. Part One – Verification Background is a description of the framework in which the verification takes place. There the mandate of the Canadian ETV Program is given, and the verification steps are defined and explained, with the aid of a block diagram. Part Two – Verification Procedure provides the protocol for examining the vendor's claim.

The Verification Organization, an expert third party, is required to follow this protocol, so that every technology developer who applies to the Canadian ETV Program is assured that there is a "level playing field", with the same procedure for every applicant.

To start the process, the technology vendor presents a comprehensive package of information for review. The Verification Organization (VO) then establishes the following:

1. The application package is complete
2. The testing agency has appropriate qualifications for conducting the test program
3. The analytical laboratories are accredited for the analysis required by this test program
4. There is sufficient technical information about the technology
5. There is credible test data, for both samples acquired for laboratory analysis and on-site process data

The VO then continues with the verification procedure:

1. After numerical analysis of the data, a performance claim is established for the technology.
2. Other elements of the performance are evaluated, thus establishing other parameters of interest for the technology. A performance claim is not associated with this review and reporting activity.
3. A report is prepared by the Verification Organization, for approval by the Canadian ETV Program and the vendor

At the conclusion of the verification, a Verification Certificate is awarded by the Canadian ETV Program. This is accompanied by the Technology Fact Sheet, highlighting the verified performance claim, and the final Verification Report.

For each step in the verification sequence, there is a checklist. These checklists are produced as a set of eleven tables. Depending on the findings, the VO will respond with Yes or No, and also with comments, some of which may be inserted directly into the checklists, which are also used as a reporting tool for the Verification Report.

After establishing the credibility of the incoming information, the data is analyzed, according to standard procedures in statistical numerical analysis. The procedures for this are covered by a series of Statistical Analysis Worksheets (SAWs). In Appendix A, there is an introduction to the basic statistical concepts required to use the SAWs, followed by a set of ten Statistical Analysis Worksheets. Further support to the VO is provided in additional appendices containing two case studies and some examples of calculations using the SAWs, as seen in Microsoft Excel spreadsheets, with associated commentary. More powerful statistical software packages may be used, where available.

Chapter 6 recognizes that the nature of some verifications do not fit the model of presentation of

data sets, followed by statistical analysis. The third party verification principles may also be applied to a vendor with certification under other jurisdictions, technology that has been recognized for conformance to science-based regulations, and technology that has not yet reached commercialization but requires the assistance of proving the principle of the innovative technology.

In summary, the technology that has met the requirements of the Canadian ETV Program General Verification Protocol, has successfully completed a rigorous examination and has achieved a high standard. The value of having a verified performance claim accrues to the vendor, the user and society as a whole.

PART ONE –VERIFICATION BACKGROUND

1.0 Purpose and Structure of Verification

1.1 Definition of Verification

Verification is a third party independent assessment and validation of a vendor's technology performance claim, following a prescribed protocol. Verification is an examination of environmental performance claims made by suppliers, and of available supporting information, for the purpose of validating the performance claims. The purpose of verification is to substantiate that the performance and integrity of the environmental technology satisfies a standardized protocol as specified by Environment Canada's ETV Program. The verification must include the confirmation, by examination and provision of objective evidence, that specified requirements are achieved. These specifications must include that an environmental product or process is based on sound scientific and engineering principles that it is effective, reliable and protective of health and environment, and that it performs in this manner under defined operating and environmental conditions.

The benefits to the vendor and stakeholders generally are:

1. The Verification Organization is an independent expert entity that has no corporate or other connection to the vendor or the Canadian ETV Program, and therefore can provide a credible third party assessment
2. By applying the General Verification Protocol, all applicants go through the same process of validation
3. The verification of the Performance Claim of the vendor supports the marketing of a new technology, generally aiding the technology to obtain widespread acceptance in the marketplace
4. Using the principles of the General Verification Protocol, a performance benchmarking program for a group of vendors with related technology could develop a technology specific test and verification protocol

1.2 The Canadian ETV Program

1.2.1 Background

The ETV Program was developed by Environment Canada in cooperation with Industry Canada and in consultation with the Canadian environment industry, and has been operating since 1998.

1.2.2 ETV Program

The Environmental Technology Verification (ETV) Program is designed to foster the growth and marketability of the technologies and processes offered by the Canadian environment industry. The Program builds on Canada's reputation by emphasizing our capabilities and credibility in the environmental market. A voluntary program, the ETV initiative has been developed to promote the commercialization of new environmental technologies into the market place. Verification gives industry the tools to provide potential buyers with the assurance that a vendor's claim(s) of performance for its environmental technology are valid, credible, based upon sound scientific and engineering principles, and supported by quality independent test data.

Environmental technology vendors apply to the ETV Program for verification of the claims they make concerning the performance of their environmental technologies. Suppliers of equipment-based environmental services (where performance can be verified) are also eligible to apply for verification.

If the claim is verified, the Canadian ETV Program issues three documents to the company:

- Verification Certificate;
- Technology Fact Sheet, and;
- Final Verification Report.

The vendor is entitled to use the ETV logo (subject to guidelines issued by the Canadian ETV Program) to market its technology in Canada and abroad. The Technology Fact Sheet, which is centred on the performance claim, is published on the Canadian ETV Program website.

1.3 Definitions

1.3.1 Environmental Technology

For the purposes of the ETV Program, environmental technologies are products and processes that offer an environmental benefit or address an environmental problem. This definition includes products and processes whose primary purpose is environmental protection or remediation. It also includes products or processes that contribute to environmentally sound production, including alternative production processes and materials. The focus is on environmental technologies and equipment-based services for industrial[1] and institutional applications.

Environmental technologies address a wide range of environmental protection and conservation needs, including:

- Pollution prevention
- Pollution detection and monitoring
- Environmentally-related human health protection
- Pollution control and treatment
- Instrumentation and measurement systems for environmental protection or remediation
- Energy efficiency/management
- Emergency response
- Non-hazardous and hazardous waste management
- Site remediation and restoration
- Land and natural resource management
- Greenhouse gas reduction/monitoring

1.3.2 Equipment – based Service

For the purposes of the Canadian ETV Program, equipment-based environmental services are services that can make claims based solely on measurable performance of the equipment or technology used. Such services can be verified in the same manner as technologies.

Excluded from consideration are "people-based" environmental services; essentially any service for which a strict performance-based verification would not be possible. Also excluded is the certification of individual environmental practitioners.

[1] The Canadian ETV Program does not deal with "green" consumer products which are addressed by Canada's Environmental Choice Program™

1.3.3 ETV Fact Sheets

Two ETV fact sheets are available. One is the Canadian ETV Program Fact Sheet which is published by the Canadian ETV Program in English, French and Spanish; it provides a general overview of the Canadian ETV Program. The second is a Technology Fact Sheet, providing vendor-specific information upon ETV Program graduation and licensing. The Technology Fact Sheet describes the verified performance claim in detail; including specific parameters, operating conditions and applications. A brief statement about the nature of the Program is also included, in addition to a statement of Limitation of Verification.

1.3.4 Verification Organizations (VOs)

Verification Organizations (VOs) are third party, impartial, specialized and accredited laboratory and testing facilities, technical review services or various technical specialists sub-contracted by the Canadian ETV Program to supply assessment and validation expertise and services. As the credibility of the vendor's verified claim is based on the data assessment and verification carried out, the VO sub-contracted by the Canadian ETV Program must have the expertise specifically relating to the technology to be verified in order to conduct the verification. Furthermore, a VO may not both generate the required data and then assess/validate that same data for any one performance claim, as this would present a conflict of interest with respect to that verification.

1.3.5 Verification Certificate

A Verification Certificate is awarded to graduates by the Canadian ETV Program upon successful validation of their performance claim. The Certificate is the vendor's authenticated proof of having successfully completed the ETV Program. It contains the graduate's full corporate/organizational identifier, description of how the technology was tested, the verified performance claim, a reference to the Technology Fact Sheet, an authorized signature by the Canadian ETV Program, a license number and effective and expiration date.

1.3.6 Verification Report

The Verification Report is issued to vendors by the Canadian ETV Program upon completion of the assessment of their performance claim. The Report contains a detailed description of the technology; a detailed description of the performance claim including specific parameters, operating conditions and applications; and the results of data assessment and claim validation.

The ETV Logo is not intended to be used alone as proof of verification, but only in conjunction with the vendor’s specific performance claim. Vendors, having successfully obtained a verification through the Program, may use the ETV Logo subject to guidelines provided by the Canadian ETV Program.

Program Requirements and Procedures

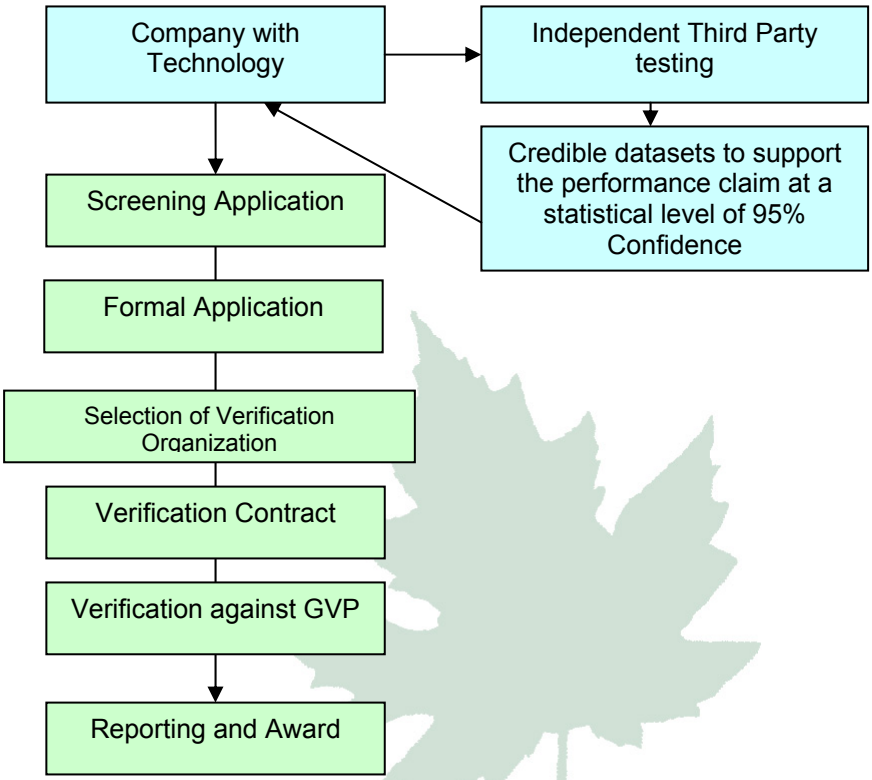
The verification of a vendor's performance claim involves the confirmation of a quantifiable claim supported by reliable data. Following a detailed and rigorous General Verification Protocol, a Verification Organization assesses the integrity of supplied data and the validity of the associated performance claim(s) based on this data. For a claim to be verified, the Canadian ETV Program must be satisfied that the following criteria have been fulfilled:

- The technology must provide a net environmental benefit.
- The technology is based on sound scientific and engineering principles.
- The claim is fully supported by independently generated, peer-review quality data, which are supplied by the applicant or generated upon the applicant's request through a test

program conducted by a qualified testing agency.

- The conditions of performance for the claim are clearly defined.

Figure 1: The Verification Process



The process of having a claim verified through the ETV Program consists of several stages.

1. In the Screening stage, for a technology to be eligible, it must be an environmental technology or an equipment-based environmental service, the performance claim must (if appropriate) meet minimum Canadian standards and/or national guidelines for that technology, and the technology must be currently commercially available or commercially ready for full-scale application. If the technology meets these criteria, the applicant submits a Screening Application to the Canadian ETV Program which is reviewed to confirm eligibility and feasibility and to resolve any conflict of interest which may exist between the applicant and the Canadian ETV Program. If the technology does not meet the criteria of commercial readiness, or if the test data for the technology is clearly insufficient, then the applicant is advised to contact the Canadian ETV Program for advice regarding test programs to generate relevant data for subsequent claim verification.

2. If the technology is eligible for application, the applicant submits a Formal Application which requires a full package of information about the technology, the claim to be verified, and the data and/or information that are currently available to support the claim. A standard non-refundable fee may apply. The Canadian ETV Program reviews the Formal Application for completeness and determines if it can be accepted into the ETV Program. If the application is not accepted, the applicant may choose to submit a modified claim, or may choose to arrange a test program for the generation of additional data. If the application is accepted, the Canadian ETV Program proposes a verification process for the claim which includes the identification of a possible Verification Organization and an estimated cost for the process.

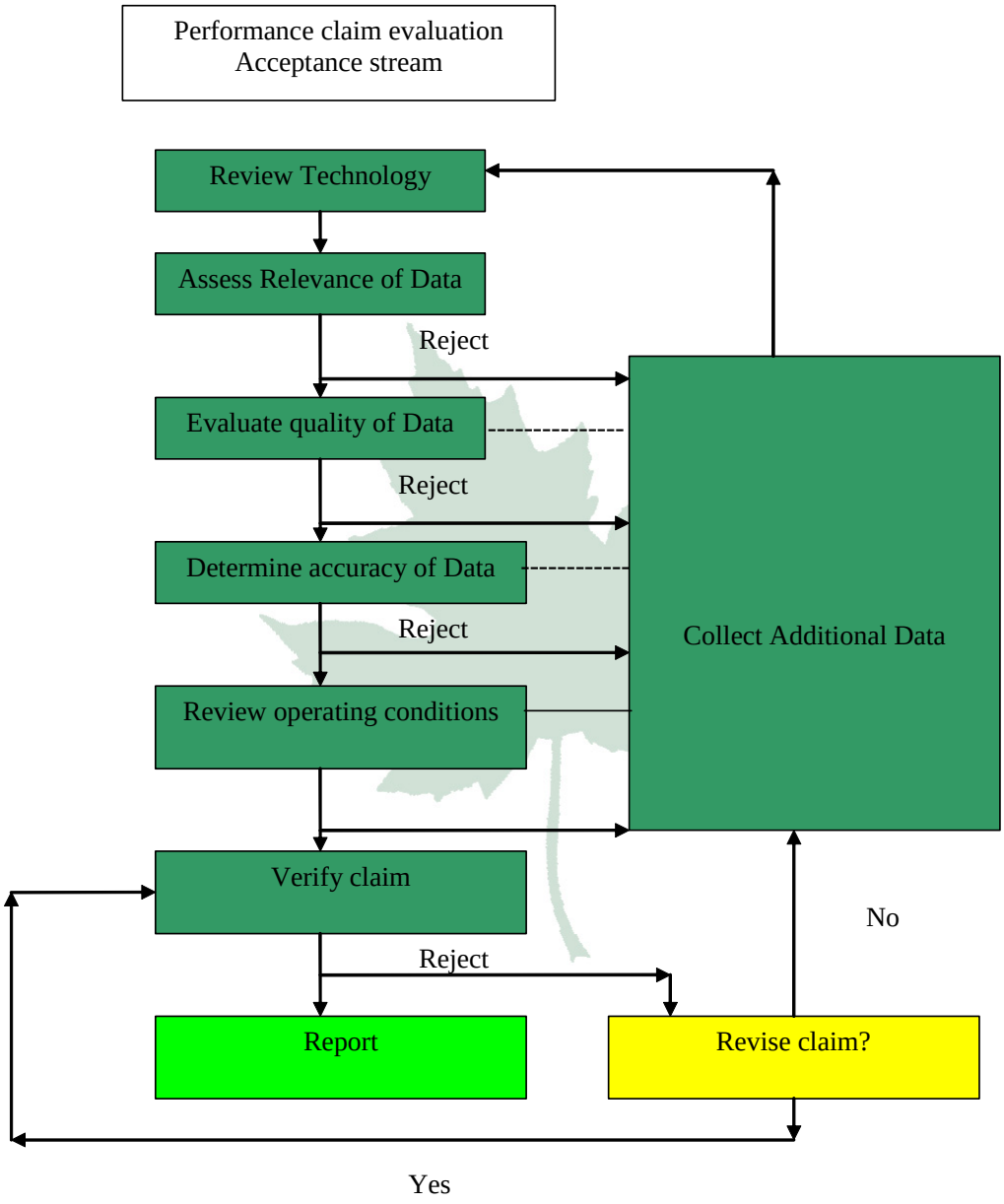
3. Before confidential information is provided to the VO or the Canadian ETV Program, Confidentiality Agreements are signed between applicant, the VO, and the Canadian ETV Program. Any conflict of interest between the applicant and the VO is resolved. The applicant enters into a contract with the Canadian ETV Program that specifies the scope and costs associated with the verification process, including administrative costs of the Canadian ETV Program.

4. During Verification, the Verification Organization follows the procedures of the General Verification Protocol. The VO reviews the complete package of information, with emphasis on the supporting data to determine if the claim is adequately substantiated or if additional data are required. The Verification Organization prepares a report on the results of the verification, and submits it to the Canadian ETV Program and the applicant for review. If the claim cannot be substantiated, the applicant may choose to modify the claim such that it may be supported by the existing data. If additional testing is required, an independent testing of the technology is conducted by an approved testing facility, or equivalent, as described elsewhere in this protocol. Costs for additional testing are paid by the applicant. The diagram below illustrates in greater detail the verification procedure at this stage.

5. At the Reporting and Award stage, where a claim has been verified, a draft Verification Report is prepared for approval by the Canadian ETV Program and the applicant, changes are made if required, and a final Verification Report is issued. The Canadian ETV Program prepares a Technology Fact Sheet to accompany the Verification Report and awards the Verification Certificate. On the third anniversary, the verification will be

up for renewal; a license renewal fee may apply. If there are no substantive changes to the technology, regulations or standards, the renewal will proceed, at the option of the Canadian ETV Program graduate.

Figure 2:



1.4 Eligibility of Applicants for the ETV Program

Several types of applicants are eligible to apply to the ETV program for verification:

1. Environmental technology vendors who have new technology.
2. Vendors who provide equipment-based environmental services that can make claims based solely on measurable performance of the equipment or technology used
3. Technology developers that have early stage technology

The Canadian ETV Program operates in Canada under a mandate from Environment Canada. However, the Canadian ETV Program welcomes applications from outside Canada, and has previously verified performance claims for technology that is not based in Canada. In addition, the Canadian ETV Program is prepared to consider data from a properly conducted test program that was not done in Canada.

Data for the Canadian ETV program will be assessed based on conformance to requirements of the General Verification Protocol. Clearly, the test plan that is used to describe the technology testing therefore needs to be sufficiently detailed to provide the test agent with the information as to correct number of and location of samples, Quality Assurance / Quality Control (QA/QC),and analytical requirements to create such data suitable for verification. Guidelines on planning the program and creating a test plan are provided by the Canadian ETV Program in the document: Guidelines to Developing a Test Program.

1.5 The Canadian ETV Program Stages

- A complete package of information is presented for review
- The testing agency has appropriate qualifications for conducting the test program
- The analytical laboratories are accredited for the analysis required by this test program
- There is sufficient technical information about the technology
- There is credible test data, for both samples acquired for laboratory analysis and on-site process data
- After numerical analysis of the data, a performance claim is verified for the technology.
- After review and /or numerical analysis, the performance is evaluated as it relates to other parameters of interest for the technology
- A report is prepared by the Verification Organization, for approval by the Canadian ETV Program and the vendor
- A Verification Certificate is awarded, accompanied by the Technology Fact Sheet, highlighting the verified performance claim, and the final Verification Report.

PART TWO – VERIFICATION PROCEDURE

2.0 Review of Application Documents

2.1 Application Documents

The Canadian ETV Program advises and works with the vendor, to ensure that there is a complete package of documents known as the Formal Application. The formal application is then delivered to the VO for a full technical review according to the procedures of the Canadian ETV Program General Verification Protocol.

The Verification Organization (VO) begins the verification process by reviewing the Formal Application form and the accompanying information and documents provided by the applicant and the Canadian ETV Program. The objective at this stage is to determine whether adequate data or information is provided and to ensure that the VO has enough information for a full understanding of the technology, the testing performed and the claims to be verified.

2.2 Application Identification

Company: _____

Product Name: _____

Technology Generic Classification and Description:

2.3 Application Review Criteria

Criteria identified in Table 1, Application Review Checklist will be reviewed in detail during the verification process. At this stage, the VO is required to check that each category of information submitted is present and appears to be complete. Note that optional items may be included with the Formal Application, and if submitted, these should be listed and assessed as to their relevance or acceptability.

Information must be supplied to satisfy the Verification Organization. All Tables, in the form of checklists are provided in electronic form to the VO, for inclusion in the Verification Report. The VO may provide short written comments directly within the table, to support a “yes” answer, and should provide justification for a “no” selection.

More extensive comments should be written near the checklist relevant to the topic. If the Verification Organization considers some criteria more important than others, strongly influencing the outcome of the verification, then the explanation for this finding should be included with the assessment and report.

Table 1: Application Review Checklist – Mandatory Information

Ref.	Criteria		Information Provided	
			Yes[2]	No
1.1	Signed Formal Application		☐	☐
1.2	Signed Declaration Regarding Codes & Standards submitted with signed formal application		☐	☐
1.3	Technology provides an environmental benefit.		☐	☐
1.4	A copy of "Claim to be Verified" for each performance claim to be verified included with the Formal Application.		☐	☐
1.5	Performance Claim composed in a way that satisfies "Criteria for Specifying Claims" :		☐	☐
	1.5.1	Include Technology name (and model number)	☐	☐
	1.5.2	Include application of the technology	☐	☐
	1.5.3	Include specific operating conditions during testing	☐	☐
	1.5.4	Does it meet the minimum requirement for the majority of Canadian Standards / Guidelines *	☐	☐
	1.5.5	Does it specify the performance achievable by the technology	☐	☐
	1.5.6	Is it the performance measurable	☐	☐
1.6	Standard operating practices and a description of operating conditions for each individual performance claim specified.		☐	☐
1.7	The proponent has supplied significant references describing or supporting scientific and engineering principles of the technology. (see Chapter 4)		☐	☐
1.8	Two or more names and contact information of independent (no vested interest in the technology) experts, qualified (backgrounds of experts are needed) to discuss scientific and engineering principles on which the technology is based. These experts must be willing to be contacted by the VO.		☐	☐
1.9	Brief summary of significant human or environmental health and safety issues associated with the technology. (Note: this criterion complements but does not replace the obligation for the applicant to submit a duly signed "Declaration Regarding Codes and Standards")		☐	☐
1.10	Brief summary of training requirements needed for safe, effective operation of technology, and a list of available documents describing these requirements. (Note: this criterion complements but does not replace the obligation for the applicant to submit a duly signed "Declaration Regarding Codes and Standards")		☐	☐

[2] Provide written justification for yes or no information provided.

Ref.	Criteria	Information Provided	
		Yes[2]	No
1.11	Process flow diagram(s), design drawings, photographs, equipment specification sheets (including response parameters and operating conditions), and/or other information identifying the unit processes or specific operating steps in the technology. If feasible, a site visit to inspect the process should be part of the technology assessment.	☐	☐
1.12	Supplemental materials (optional) have been supplied which offer additional insight into the technology application integrity and performance, including one or more of :		
	A copy of patent(s) for the technology, patent pending or submitted.	☐	☐
	User manual(s).	☐	☐
	Maintenance manuals.	☐	☐
	Operator manuals.	☐	☐
	Quality assurance procedures.	☐	☐
	Sensor/monitor calibration program.	☐	☐
	Certification for ISO 9001, ISO 14000, or similar program.	☐	☐
	Material Safety Data Sheet (MSDS) information.	☐	☐
	Workplace Hazardous Materials Information System (WHMIS) information.	☐	☐
	Health and Safety plan.	☐	☐
	Emergency response plan.	☐	☐
	Protective equipment identified.	☐	☐
	Technical brochures.	☐	☐
1.13	The applicant provided adequate documentation and data. There is sufficient information on the technology and performance claim for the performance claim verification. [If necessary, the VO should communicate with the Canadian ETV Program to request copies of the necessary documentation and required data that are available to support the claims.]	☐	☐

2.4 Qualifications and Independence of Test Agency and Analytical Laboratory

The goal of verification is to establish credibility of the claim of the vendor. Thus, the starting point is at the time the testing, data collection and sample analysis takes place. The testing program requires third party involvement, either through a test agency, or participation in the tests carried out by the vendor, by a qualified individual expert who is independent of the vendor.

2.4.1 Qualifications and Independence of Test Agent

Participation of either a test agency, or an independent expert participant is mandatory, and for the purposes of this protocol, either will be designated by the terminology “test agency”. It is advisable that the vendor’s choice of an independent, unbiased test agency should be done in consultation with the Canadian ETV Program. The (pre-qualified) test agency shall review this

test protocol in detail and conduct the verification testing according to the test plan. It is the testing agency's responsibility to operate and maintain the technology in accordance with the vendor's O & M manual. If necessary, the testing organization can modify and make changes to the existing test plan. Changes should be documented and undertaken in consultation with the Canadian ETV Program before testing is implemented. Upon completion of the test, the testing agency should prepare and submit a Testing Report, including the test data in its original form, i.e. not consolidated, averaged, or otherwise different from the "raw" data. The Testing Report is submitted to the owner of the technology, who then submits it to the Canadian ETV Program.

2.4.2 Qualifications and Independence of Analytical Laboratory

It is a mandatory requirement that samples be submitted to an accredited laboratory, one that has been certified for analyzing specific parameters by the Canadian Association for Environment and Analytical Laboratories (CAEAL) and is accredited to ISO 17025 - General requirements for the competence of testing and calibration laboratories. The laboratory shall have a well developed Quality Assurance (QA) plan. It is the laboratory's responsibility to apply and execute the appropriate analytical procedures which meet general accepted principals of good laboratory practice and quality control. Appropriate laboratory equipment shall be provided for sample analysis. Chains of custody and records of analytical procedures must be maintained throughout the process. It is helpful if the laboratory has related experience with similar projects.

2.4.3 Test Programs that do not involve an Analytical Laboratory

Depending on the nature of the technology, the measurements may not require the services of an analytical laboratory. In the test plan, this information should be given, in the section describing the methodology of data acquisition. The primary objective of every test plan should be to generate credible data and so the proponent should explain the plan for achieving this objective.

2.5 Technology Type: Specific Verification Protocol & Supplemental Documents

The Canadian ETV Program's General Verification Protocol (GVP) does not cover every possible technology or process that may require performance verification. The GVP outlines the principles of verification, so when a technology-specific verification protocol must be written, the basic framework may be acquired from the GVP. If supplemental documents are required, the vendor, the Canadian ETV Program and the VO will discuss this.

2.6 Review of Conformance to Performance Benchmarking

In addition to technology specific verifications, the Canadian ETV Program provides sector- and program-based performance benchmarking services. These services address the need to develop acceptable performance criteria which can be used for a group of stakeholders who have a common technology requirement.

Groups such as provincial or national trade associations may work with the Canadian ETV Program to develop alternate and/or additional requirements for verification of performance of a technology type that many of its members are considering purchasing. For example, fleet managers may find many vendors offering emission reduction equipment to install on existing vehicles.

3.0 Review of Technology

3.1 Scientific and Technical Foundation

Each verification must include evidence that the technology to be verified is based on sound scientific and technical/engineering principles. The Verification Organization (VO) must confirm that information has been provided or referenced ensures a clear understanding of the candidate technology, including the scientific and technical principles of operation.

Peer reviewed scientific literature is generally accepted as the highest quality form of reference and, if available, should be supplied or referenced with the application. Another source may be documentation in reports written for review by technical experts. Technical principles may be explained in detail by the proponent, or by reference to a textbook. If the textbook is not widely available, the proponent may be required by the VO to provide a copy of it. The purpose is to convey the essential background information for the reviewer to fully understand the technology to be reviewed. In addition, the proponent should provide a reference list of technical articles and relevant regulations or standards that are pertinent to the performance claim. The VO should confirm that the information supplied is complete and correct, as determined on the basis of best professional judgment.

It will be necessary for the VO to read the key articles, view process flow diagrams, equipment specification sheets, etc. supplied by the proponent, and the other citations listed in the application (e.g., technical papers, textbooks, etc.). In some cases, it may be advantageous to contact an independent expert to obtain additional information. The VO may also contact the vendor's references, including customer references if available. Draft documents, reports in print, and other text may also prove useful in determining the scientific principles of a technology, but do not carry the substance or regard of peer review quality literature. For example, in the case of a treatment technology, the applicant/proponent may include balanced measures of inputs and outputs, energy and water, and process flow diagrams. In summary, selected reference and technical material must be provided by the proponent, so that it may be used directly and indirectly to support the claim.

A situation where scientific and technical principles may not be evident is one where an established technology has known performance and operation under conditions of substantiated theory, but is newly extended under conditions that require the introduction of innovative scientific or technical principles. Although a certain performance has been characterized by an experimental test program, the underlying theory may not be well understood. Another example is a case where energy or mass balances are not obtainable. In these cases, the VO must use best professional judgment, in consultation with the Canadian ETV Program.

If the VO has any questions regarding the scientific basis of the technology, they must either initiate a discussion with the proponent or have the Canadian ETV Program facilitate a request for additional information. This should be done as early in the verification process as possible. If the VO does not feel there is sufficient documentation and evidence to explain the scientific principles, then the VO should include in the final report an explanation of the basis for verification of the claim, and also describe the basic scientific and technical principles that are not sufficiently described by the proponent or understood by the VO. The VO must use best professional judgment in this situation.

In cases where the scientific and technical principles are not well understood, it is essential to

the decision making process that the data and the conditions under which it was acquired (e.g. proper sampling and sample transport procedures) be seen to be of high integrity and inherent quality and reproducibility. Data validity could be major factors in the VO's expert opinion regarding whether the technology is based on sound scientific principles.

The validity of the data (further discussed in Chapter 4) can be established based on the following criteria:

- Existence of sufficient baseline data for comparative purposes, complete and representative of the baseline case
- Reliable operational data
- Demonstration of the impact of any input or process variability
- Appropriate experimental design and proper test conditions – proper sampling protocol and analytical procedures; calibrated testing equipment
- Technology operations monitored and recorded during testing
- Quality Assurance and Quality Control procedures
- Full documentation and complete chains of custody
- Interviews (with vendor references): A brief report should be written if the VO has interviewed reference contacts given by the vendor.

3.2 Operation and Maintenance

Operation of the units verified is usually only based on the operating conditions of the technology during the testing period. Verification does not involve maintenance issues directly as these are not usually part of the testing.

3.3 Technology Review Criteria

Table 2 must be completed for each environmental technology performance claim (or group of claims). If the Verification Organization considers some criteria especially important, or has other comments, this information must be documented and included with the assessment and report. Short comments may be included directly within the checklist text, and questions that are not applicable should be so noted.

Table 2: Technology Review Criteria Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[3]
Technology Description			
2.1	Technology based on scientific and technical principles. (It will be necessary for the VO to read the key articles and citations listed in the Formal Application. It may also be necessary to contact the independent experts listed in the Formal Application to obtain additional information.)	☐	☐

[3] Provide written justification for no meets criteria.

Ref.	Criteria	Meets Criteria	
		Yes	No[3]
2.2	Technology supported by peer review technical literature or references. (Peer review literature and texts must be supplied with the Formal Application as well as relevant regulations and standards that are pertinent to the performance claim)	☐	☐
2.3	Technology designed, manufactured, and/or operated reliably. (historical data from the applicant, not conforming to all data criteria, may be useful for the VO to review to assess the viability of the technology not for verification, but for insight purposes)[4]	☐	☐
2.4	Technology designed to provide an environmental benefit and not create an alternative environmental issue. (e.g. it does not create a more hazardous and or unmanaged byproduct and it does not result in the transfer of an environmental problem from one media to another media without appropriate management of the subsequent contaminated media)	☐	☐
2.5	Technology conforms to standards for health and safety of workers and the public.[5] The vendor must submit a signed "Declaration Regarding Codes & Standards", with the Formal Application. The role of the Verification Organization is to ensure this signed document is included with the information that is reviewed for the performance claim verification	☐	☐

[4] Also note **The** VO **should use best judgment and apply standards relevant to the technology sector to generally assess whether the technology has been designed and manufactured in an acceptable fashion. A critical assessment of the materials / apparatus used in the technology is beyond the scope of the ETV program. Any assessment of the integrity of the manufacture of technology components must be performed by personnel whose experience and expertise qualify them to undertake this activity. It is not the responsibility of the Verification Organization to assess the integrity of materials and substances used in the manufacture of the technology, other than to understand their use and implication on the performance of the technology.**

It is the vendor's responsibility to ensure that applicable regulations and guidelines are satisfied with respect to application of the technology. The vendor must submit a signed "Declaration Regarding Codes & Standards", generally with the Formal Application. The role of the Verification Organization is to ensure this signed document is included with the information that is reviewed for the performance claim verification.

Claim verification by the Verification Organization does not represent any guarantee of the performance or safety of the equipment or process. The Verification Organization shall not be liable in any way in the event that the device or process fails to perform as advertised by the supplier or as expected by the consumer. The Verification Organization shall not be liable for any injury to person or property resulting from the use of the equipment or process.

[5] For the purposes of the ETV Program, the health and safety issue has been defined as a subjective criteria, requiring a value judgment on the part of the reviewer as to the integrity or reliability of any or all health and safety documentation provided by the applicant. As such, the Verification Organization cannot assume any liability in making a "Best Professional Judgment" assessment of the technology using these

Ref.	Criteria	Meets Criteria	
		Yes	No[3]
Environmental Standards			
2.6	Technology achieves federal, provincial, and/or municipal regulations or guidelines for management of contaminated and or treated soils, sediments, sludges, or other solid-phase materials.	☐	☐
2.7	Technology achieves federal, provincial, and/or municipal regulations or guidelines for all (contaminated and or treated) aqueous discharges as determined by the applicants information.	☐	☐
2.8	Technology achieves federal, provincial, and/or municipal regulations or guidelines for all (direct or indirect) air emissions. If the environmental technology results in the transfer of contaminants directly or indirectly to the atmosphere, then, where required, all regulations or guidelines (at any level of government) relating to the management of air emissions must be satisfied by the applicant's information.	☐	☐
Commercial Readiness			
2.9	Technology and all components (apparatus, processes, products) is full-scale, commercially-available, or alternatively see 2.10 or 2.11, and, data supplied to the Verification Organization is from the use or demonstration of a commercial unit.	☐	☐

criteria.

(continued footnote 5 from Ref. 2.5) A critical validation of the Health and Safety aspects of the vendor's technology is beyond the scope of the ETV program. Any validation of health and safety issues must be performed by personnel whose experience and expertise qualify them to undertake these activities. Staff from noted organizations and agencies [e.g., Health and Welfare Canada (H&W), Provincial Labour Ministries, Industrial Accident Prevention Association (IAPA), [US] Occupational Safety and Health Association (OSHA), water pollution control agencies, province/state health departments, fire protection associations, etc.], may be able to provide advice or technical services on these issues. It is **NOT** the responsibility of the Verification Organization to validate the Health and Safety aspects of the technology.

It is the vendor's responsibility to ensure that regulations and guidelines are satisfied in the application of the technology. The Verification Organization can request additional written confirmation from the applicant that the company has sufficient documentation to address worker health and safety issues and requirements related to the use of the technology, including an Emergency Response Plan.

Ref.	Criteria	Meets Criteria	
		Yes	No[3]
2.10	Technology is a final prototype design prior to manufacture or supply of commercial units, or alternatively see 2.11, Note: Verification of the performance claim for the technology is valid if based on a prototype unit, if that prototype is the final design and represents a pre-commercial unit. The verification will apply to any subsequent commercial unit that is based on the prototype unit design. The verification will not be valid for any commercial unit that includes any technology design change from the prototype unit used to generate the supporting data for the verification.	☐	☐
2.11	Technology is a pilot scale unit used to provide data which when used with demonstrated scale up factors, proves that the commercial unit satisfies the performance claim.[6]	☐	☐
Operating Conditions			
2.12	All operating conditions affecting technology performance and the performance claim have been identified.	☐	☐
2.13	The relationships among operating conditions and their impacts on technology performance have been identified. Note: It is the responsibility of the VO to understand the relationship between the operating conditions and the performance of the technology, and to ensure that the impacts of the operating conditions and the responses of the technology are compatible.	☐	☐
2.14	Technology designed to respond predictably when operated at normal conditions (i.e. conditions given in 2.12), and/or alternatively see 2.15, Note: The Verification Organization must be satisfied that these data do not demonstrate a performance that is different than the performance indicated in the Performance Claim to be validated.	☐	☐
2.15	Effects of variable operating conditions, including start up and shut down, are important to the performance of the technology and have been described completely as a qualifier to the performance claim under assessment.	☐	☐

[6] In exceptional situations, data from a pilot scale unit may be used to validate a performance claim. This situation can be permitted if the pilot scale unit is a "scaled down" model of a full size commercial unit and engineering scale-up factors have been provided by the applicant as part of the verification process. The performance claim verification must include validating the scale-up factors.

Ref.	Criteria	Meets Criteria	
		Yes	No[3]
Throughput Parameters			
2.16	Effects of variable contaminant loading or throughput rate must be assessed and input/output limits established for the technology. Note: If the application of the technology is to a variable waste source or expected (designed) variable operating conditions, then it will be necessary to establish acceptable upper and lower ranges for the operating conditions, applications and/or technology responses. Sufficient, quality data must be supplied to validate the performance of the technology at the upper and lower ranges for the operating conditions, applications and or technology responses detailed in the performance claim.	☐	☐
Other Relevant Parameters/Variables/Operating Conditions The Verification Organization is expected to understand the technology and identify and record all relevant criteria, parameters, variables or operating conditions that potentially can or will affect the performance of the technology under assessment. It is practical to include all of these variables in Table 2 (from 2.17 to …).			
2.17		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
2…	Continue on attached page(s) as required.	☐	☐

2 Provide written justification for yes or no meets criteria.

4.0 Review of Test Plan, Test Execution and Data

4.1 Review of Test Plan and Execution of Test Plan

Compare the activities performed during the testing to the original test plan. When the execution of the Test Plan differs from the original plan, the differences should be noted and the effect on the performance claim(s) should be explained.

4.2 Review of Original Data

The performance claim(s) should be supported by peer-reviewed third party data. Data quality requirements are listed below:

The evaluation of analytical data involves more than a review of summary results from analytical tests. The Verification Organization will assess suitability of the data using the following criteria:

1. Data suitable for testing the performance claim.
2. Samples representative of process characteristics at the specified locations.
3. Samples representative of testing conditions.
4. Samples representative of appropriate operating conditions.
5. Adequate number of samples.
6. Samples and data prepared by a third party independent test agent or through extensive on site observation of the vendor conducted test program by an independent test agent.

The quality of the analytical data provided will be assessed using the following criteria:

7. Appropriate sample collection methods.
8. Apparatus and/or facilities for the test adequate for the generation of data.
9. Operating conditions during test adequately monitored and documented.
10. Operating conditions and measuring equipment measured/calibrated at sufficient frequency.
11. Acceptable QA/QC procedures followed during sample collection.
12. Chain-of-custody methodology used for sample handling and analysis.

This step in the verification process involves a review of the verification study design, data validity and acceptability concerning the specific technology performance claim(s) being made. The objective at this stage is to ensure that the technology data set meets the verification criteria.

4.2.1 Data Set Identification

It is essential for all the data used in the verification that (at least) the following information pieces are included.

Title of Data Set: [concise and easily understandable]

_____

Date of Data Set: _____

Test agent

_____

Test site:

_____

Test Laboratory

4.3 Review of Documentation from Other Institutions

The verification data set submitted in the Formal Application from the client may be comprised of data sets from other Institutions. However, such Institutions need to be accredited for the tests performed. Evidence of no conflict of interest between the Institution and the proponent as well as submission of the Institutions official testing reports is essential as is chain of custody information/forms where appropriate. Raw data (where possible) may also be required in a review.

Table 3: Verification Study Design Assessment Criteria Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[7]
3.1	Was a statistician, or an expert with specialized capabilities in the design of experiments, consulted prior to the completion of the test program, and if so please provide the contact details.[8]	☐	☐
3.2	Is a statistically testable hypothesis or hypotheses provided? (so that an objective, specific test is possible)[9]	☐	☐
3.3a-c	Does the verification study generate data suitable for testing the hypothesis being postulated? [10] Namely:	☐	☐
3.3a	Does the study measure the parameters used in the performance claim hypothesis?	☐	☐
3.3b	Does the study control for extraneous variability?	☐	☐
3.3c	Does the study include only those effects attributable to the environmental technology being evaluated?	☐	☐
3.4	Does the verification study generate data suitable for analysis using the SAWs? (i.e. it is preferable that tests are designed with the SAWS in mind before test plans are written)	☐	☐
3.5	Does the verification study generate data suitable for analysis using other generic experimental designs (ANOVA etc)? (clearly, verification studies should be designed with the final data analysis in mind to facilitate interpretation and reduce costs)	☐	☐

[7] Provide written justification for yes or no meets criteria.

[8] An expert statistician can help determine during the experimental design which experimental variables need to be controlled and or monitored so as to be able to defend a verification claim

[9] The hypothesis that Statistical Analysis Worksheets will test are of the general form:

What is the degree of confidence around a measured result?
Is a mean equal to a specified value?
Is a median equal to a specific value?
Is mean 1 = mean 2 ?
Is median 1 = median 2 ?
Is variance 1 = variance 2 ?
Can a process change an influent/product/waste by ‘p’ percent?
Are two paired measurements different?

[10] Note: When data are not available on a specific parameter, it may be possible to use data on a surrogate parameter that has known correlation to the unmeasured parameter. In this case, the correlation must be clearly defined, demonstrated and based on sound scientific, engineering and or mathematical principles. The applicant must submit that data for their set of tests.

3.6	Are the appropriate parameters, specific to the technology and performance claim, measured? (it is essential that the VE and the technology developer ensure that all parameters – e.g. temperature etc - that could affect the performance evaluation are either restricted to pre-specified operating conditions or are measured)	☐	☐
3.7a-d	Are samples representative of process characteristics at specified locations?. namely:		
3.7a	Are samples collected in a manner that they are representative of typical process characteristics at the sampling locations for example the samples are collected from the source stream fully mixed etc	☐	☐
3.7b	Is data representative of the current technology?	☐	☐
3.7c	Have samples been collected after a sufficient period of time for the process to stabilize?	☐	☐
3.7d	Have samples been collected over a sufficient period of time to ensure that the samples are representative of process performance?	☐	☐
3.8	Are samples representative of operating conditions? Note: A time lag occurs between establishing steady state conditions and stabilization of the observed process performance. This time lag depends in part on the time scale of the process. (i.e. for a Completely Stirred Tank Reactor (CSTR) flow-through system, the time scale is determined by the residence time of the contaminants in the reactor. It is usual that at least three residence times are required to achieve effective stabilization. Therefore if sampling has been performed from a CSTR, then sampling should have only begun after at least three hydraulic residence times had occurred, and testing continued for at least an additional three residence times to ensure that the aggregate data set is representative of process performance)	☐	☐
3.9	Are samples representative of known, measured and appropriate operating conditions? (Note: this includes technologies that operate on short cycles and so have start and stop cycles which affects the operation of the technology). If the operating conditions are not vital but are recommended, then the reviewer must evaluate operating conditions,	☐	☐
3.10	Were samples and data prepared or provided by a third party? (Note: In some cases, where the expertise rests with the applicant, an independent unbiased third party should witness and audit the collection of information and data about the technology. The witness auditor must not have any vested interest in the technology.)	☐	☐
3.11a-c	Verification Study Design is Acceptable Namely:		

3.11a	The samples have been collected when the technology was operated under controlled and monitored conditions.	☐	☐
3.11b	A verification study design should have been established prior to the test to ensure that the data were collected using a systematic and rational approach	☐	☐
3.11c	Verification Study Design should have defined the acceptable values or ranges of values for key operating conditions, and the data collection and analysis methodology	☐	☐

4.4 Data Validity Checklist

The data validity checklist criteria help the VO determines whether a datum represents the conditions described in the performance claim. The data validity checklist also ascertains whether or not samples have been collected, transported and analyzed in a manner that does not introduce undue extraneous variability.

Table 4a: Data Validity Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[11]
4.1	Were appropriate sample collection methods used (e.g. random, judgmental, systematic etc?). For example: simple grab samples are appropriate if the process characteristics at a sampling location remain constant over time. Composites of aliquots instead may be suitable for flows with fluctuating process characteristics at a sampling location. Note: Sampling methods appropriate for specific processes may sometimes be described in federal, provincial or local monitoring regulations	☐	☐
4.2	Were apparatus and/or facilities for the test(s) adequate for generation of relevant data? (i.e. testing was performed at a location and under operating conditions and environmental conditions for which the performance claim has been defined.)	☐	☐
4.3	Were operating conditions during the test monitored and documented and provided?	☐	☐
4.4	Has the information and or data on operating conditions and measuring equipment measurements and calibrations been supplied to the Verification Organization?	☐	☐
4.5	Were acceptable protocols used for sample collection, preservation and transport (acceptable protocols include those developed by a recognized authority in environmental testing such as a provincial regulatory body, ASTM, USEPA, Standard Methods)?	☐	☐

[11] Provide written explanations for yes or no meets criteria.

4.6	Were Quality Assurance/Quality Control (QA/QC) (e.g. use of field blanks, standards, replicates, spikes etc) procedures followed during sample collection? A formal QA/QC program, although highly desirable, is not essential, if it has been demonstrated by the vendor's information that quality assurance has been applied to the data generation and collection.	☐	☐
4.7	Were samples analyzed using approved analytical protocols? (e.g. samples analyzed using a protocol recognized by an authority in environmental testing such as Standard Methods, EPA. ASTM etc. Were the chemical analyses at the site in conformance with the SOPs (Standard Operating Procedures) ?	☐	☐
4.8	Were samples analysed within recommended analysis times (especially for time sensitive analysis such as bacteria)	☐	☐
4.9 a-e	Were QA/QC procedures followed during sample analysis Including?	☐	☐
4.9a	Maintaining control charts	☐	☐
4.9b	Establishing minimum detection limits,	☐	☐
4.9c	Establishing recovery values	☐	☐
4.9d	Determining precision for analytical results	☐	☐
4.9e	Determining accuracy for analytical results	☐	☐
4.10 a-c	Was a chain-of-custody (full tracing of the sample from collection to analysis) methodology used for sample handling and analysis. Namely:	☐	☐
4.10a	Are completed and signed chain-of-custody forms used for each sample submitted from the field to the analytical lab provided for inspection to the Verification Organization?	☐	☐
4.10b	Are completed and easily readable field logbooks available for the VO to inspect?	☐	☐
4.10c	Are their other chain-of-custody methodology actions and documentation recorded/available (e.g. sample labels, sample seals, sample submission sheet, sample receipt log and assignment for analysis)	☐	☐
4.11	Experimental Data Set is Acceptable (the quality of the data submitted is established using the best professional judgment of the VO)	☐	☐

Remote monitoring data (e.g. telemetry)
For Data produced at sites which need the data to be sent electronically from the on-site instrument to the data receiving site (node) needs to be assessed for its integrity. Table 4b details the areas of the data integrity needing assessment by the VO.

Table 4b: Remote Sensing Data

Ref.	Criteria	Meets Criteria	
		Yes	No[12]
4.12	Does the remote data device (e.g. meter) authenticate all sending and receiving nodes prior to any data transfer	☐	☐
4.13	Is data sent from remote monitoring device encrypted during transfer	☐	☐
4.14	Is the data received with 100% integrity	☐	☐
4.15	What methods can be demonstrated that the data is received with 100% accuracy	☐	☐
4.16	Experimental Remote Data Set best practices are Acceptable (the quality of the data submitted is established using the best professional judgment of the VO)	☐	☐

Note, depending on the nature of the verification, the checklists may need to be modified and or new questions developed based on discussion with the Canadian ETV Program and the vendor.

5.0 STATISTICAL ANALYSIS

The Canadian ETV Program verifications require in most cases that the test data shows that the performance claim is statistically significant to 95%. As a result, the data collected from the testing of the technology needs to be analyzed with various Statistical Analysis Worksheets (SAWS). These SAWS are contained as Excel spreadsheets from the Canadian ETV Program (Appendix A) which describe examples of statistical analysis on test program data.

5.1 Formation of a Hypothesis Regarding Performance

5.1.1 The Null Hypothesis

When sampling from a population, it is not possible to prove that a performance claim or hypothesis is true. We can only disprove hypotheses and accept an alternative hypothesis or performance claim. Therefore we must pose a performance claim in such a way that the process of disproving it, verifies the claim. The hypothesis being disapproved is referred to as a null hypothesis and the hypothesis being accepted, after rejecting the null hypothesis is the alternative hypothesis. These are designated as Ho and Ha, respectively. An example will serve to clarify these concepts.

In the previous example, the performance claim was that the process generating the effluent would result in a BOD of 240 mg/l or less. An example null hypothesis Ho is:

The process produces an effluent with a mean BOD equal to 240 mg/l,

or more succinctly as:

[12] Provide written explanations for yes or no meets criteria.

Ho: μBOD=240 mg/l.

The alternative hypothesis (what we are interested in accepting, not proving) may be phrased as:

The process produces an effluent with a mean BOD less than 240 mg/l,

or more succinctly as:

Ha: μBOD = 240 mg/l.

If we can reject the null hypothesis we can accept the alternative hypothesis and therefore verify the performance claim.

5.1.2 Data Assessment

The performance claim to be verified should be stated as hypotheses that may be objectively evaluated using appropriate statistical methodology (for example, the Statistical Analysis Worksheets). The checklists in Chapter 5 are to be used to determine if the design of the verification study to verify performance claims is appropriate to the hypothesis(es) being tested. These criteria address the appropriateness of the process samples that generated the data. Justification for a "yes" or "no" designation in the written report should be provided by the VO.

It is understood that requiring large sample numbers that dictates extensive time, effort, cost could make the test economically and practically unfeasible. Many standards require results of only one sample and are the accepted norm especially if test conditions and measurements are carefully designed and controlled, even though they would not pass a statistical test. Many years of successful standards history have confirmed that this is an acceptable process. The vendor and the Canadian ETV Program should discuss these issues.

5.2 Degree of Confidence and Confidence Intervals

We may want to ascertain the degree of confidence regarding some performance characteristic of an environmental technology. To introduce the idea of the degree of confidence, we first introduce the "confidence interval". We could ask the question: For what range of values would we be 95% certain that the range contains my true process characteristic? The process characteristic might be the mean BOD.

Notice that the range of values falls around the true process characteristic not the sample value. We do not know what the value of the true process characteristic is. We conduct a verification experiment to estimate this value. This estimate is only based upon a sample of the possible outcomes (not all of them). Obviously, as we increase the number of samples so that we are sampling a larger proportion of all possible outcomes we are more and more certain of our estimate. If we sample the entire population, there is no uncertainty about our estimate. The difference between a sample and a population is described in the glossary.

By assuming that our sample estimate arises from a given statistical distribution we can determine the range of values for which we could say: "we are 95% certain that the unknown process characteristic lies within these two values".

This is known as a confidence limit. If we wish to be very certain about an estimate we can

increase the number of samples and reduce the size of the confidence interval. Similarly, if the measurements used to estimate the statistic of interest are extremely variable, we are less confident about our estimate. The confidence interval reflects this; a highly variable sample will produce a larger confidence interval than a less variable sample. The size of the confidence interval may be reduced by increasing the sample size or controlling sources of variability through the experimental design and/or analysis.

5.3 Statistical Analysis of the Performance Claim(s)

This step in the verification process involves the statistical analysis review of the performance claim(s). The Statistical Analysis Worksheets (SAWs) contained in Appendix A may be used to mathematically evaluate the performance claim(s).

5.3.1 Performance Claim

The vendor is required to give a preliminary Performance Claim with the Formal Application. This will be used initially to formulate a hypothesis. However, the proven Performance Claim may differ from that suggested by the vendor.

5.3.2 Performance Claim(s) Verification

The verification of each technology performance claim(s) requires application of the Statistical Analysis Worksheets (SAWs) to all data sets that were rated as satisfactory from the data assessment process.

The data set(s) provided to support the performance claim should be evaluated using the Statistical Analysis Worksheets in Appendix A. The SAWs were chosen to provide analytical methods for the most common types of data sets generated by verification experiments. They are suitable for use by the non-statistician, provided test assumptions are verified and the concepts emphasized in the GVP and SAWs are understood and used when data interpretations are made.

5.4 Data Analysis Checklist

The intent of the data analysis checklist is to ensure that the appropriate statistical tools can be used in a rigorous, defensible manner.

Table 5: Data Analysis Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No2
5.1	Does the analysis test the performance claim being postulated? (When conducting performance evaluations, under the ETV program, the alternative hypothesis of a "significant difference" without stating the direction of the expected difference will usually be unacceptable)	☐	☐
5.2	Does the analysis fit into a generic verification study design? (Many other "generic" designs exist that are not explicitly covered by the ETV program (e.g. ANOVA, ANCOVA, regression etc) that are potentially useful)[13]	☐	☐
5.2 a-c	Are the assumptions of the analysis met. Namely: (a negative response to 3.30 a-c means the VE needs to request further information)		
5.2.a	Did the data analyst check the assumptions of the statistical test used?	☐	☐
5.2.b	Are the tests of assumptions presented?	☐	☐
5.2.c	Do the tests of the assumptions validate the use of the test and hence the validity of the inferences?	☐	☐
5.3	Data Analysis is Acceptable The data analysis is acceptable if the statistical test employed tests the hypothesis being postulated by the technology developer, the assumptions of the statistical test is met and the test is performed correctly.	☐	☐

5.5 Data Interpretation Checklist

The intent of the data interpretation checklist is to ensure that the data analyses results are interpreted in a rigorous, defensible manner. The checklist also emphasizes that an initial performance claim may be rewritten and updated to better reflect what the data support, using the expertise of the VO and other pertinent resources.

Table 6: Data Interpretation Checklist

[13] Examples of potentially useful verification study designs or analyses not covered by the ETV program are:

- completely randomized designs with more than two treatments (ANOVA);
- designs where some of the operating conditions vary widely enough to require acknowledgement both in the experimental design and analysis stage. (ANCOVA, regression);
- analysis of count data such as microbial counts; and,
- analysis of proportional data such as proportion of organisms responding to a treatment.

Ref.	Criteria	Meets Criteria	
		Yes	No[14]
6.1a	Are the results statistically[15] or operationally significant? Did the verification result in a statistically significant test of hypothesis?	☐	☐
6.1b	To be operationally significant, does the technology meet regulatory guidelines and applicable laws?	☐	☐
6.2	Does the verification study have sufficient power to support the claim being made? Note: For verification study designs where acceptance of the null hypothesis results in a performance claim being met, the statistical power of the verification study must be determined A statistical power of at least 0.8 is the target. If the power of the verification experiment is less than this value the VO should contact the Canadian ETV Program to discuss an appropriate course of action. See Appendix A for examples on calculating sample size	☐	☐
6.3	Is the interpretation phrased in a defensible manner? Note: The final performance claim should reflect any changes to the claim made during the course of the analyses, variations or restrictions on operating conditions, etc. that changed the scope of the performance claim. The initial performance claim should be viewed as a tentative claim that is subject to modification as the verification progresses. A thoughtful open-minded verification will in the end, prove to be of greatest benefit to the technology developer.	☐	☐
6.4	Data Interpretation is Acceptable The data interpretation is acceptable if the data analyses results are reviewed in a manner that emphasizes the applicability to the specific performance claim and the statistical power of the verification experiment.	☐	☐

In the case where an understanding of otherwise unsubstantiated theory or principle may benefit from a further test program to produce an extended data set, gap analysis may be undertaken

[14] Provide written justification for yes or no meets criteria.

[15] In some cases, a new statistical approach may be necessary in order to analyze the data provided. If the existing Statistical Analysis Worksheets (SAWs) provided in the General Verification Protocol (GVP) do not apply, any other proposed approaches should be discussed with and approved by the Canadian ETV Program. In these cases, the preferred course would be to have additional SAWs developed by the Canadian ETV Program.

by the VO to determine what additional data is required to verify the proposed claim. If this can be determined early in the verification process, this feedback can be provided to the proponent so that they may provide the additional data in a timely manner. In this situation, it may be necessary for the original performance claim to be revised after the VO has reviewed and analyzed the additional data as submitted.

If the Verification Organization feels that the technology is not based on known scientific and technical principles although the data sets submitted are complete and the statistical analysis substantiates the performance claim, there may be a need to take the approach of disproving that the technology performs as claimed. In this case, the data provided with the claim proves that the technology functions as stated and there is no other cause or effect evident to prove otherwise. The explanation for the performance of the technology would then be the existence of the components and processes that have been tested and analyzed. This approach must ultimately be the VO's decision, facilitated through ETV. A thorough documentation of the process is required and is to be a component of the final verification report. If the proponent does not agree with the VO's decision, they have the option to request a second opinion by an alternate VO. Any costs incurred as a result of this shall be paid by the proponent.

5.6 Summary of Acceptable Datasets for Verification

A summary of the statistical analyses, highlighting the data sets and the specific statistical analysis worksheets used, can be summarized in Table 7 below.

Table 7: Summary of Acceptable Data Sets for Verification

Acceptable Data Set(s) Identification	SAWs Used[16]	Support Claim	
		Yes	No
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐

If a claim cannot be verified by the available data, the applicant has three options:

13. **Modify the claim** to suit the available data (see Revision of Performance Claims).

14. **Generate additional data** by having new tests conducted. The additional data would be treated as a new data set that would be evaluated using the General Verification Protocol.

15. Withdraw from the ETV Program. A summary report is to be prepared on the data assessment and claim verification processes as described in the section on Report Preparation.

5.7 Revision of Performance Claims

[16] Refer to Appendix A.

If an assessment of the available data indicates that it is acceptable for subjecting it to the performance claim verification process, but some or all of the claims cannot be verified by these data, the applicant may wish to modify the non-verifiable performance claims to suit the available data.

Completed details of any revisions to the performance claim must be provided by the VO in the final verification report, but do not need to be expressed on the verification certificate.

6.0 Alternate Means of Establishing Verified Performance

Alternate means of establishing verified performance can possibly be acceptable for the Canadian ETV Program verification use. These include Evidence of Proof, sufficient numerical analysis and testing reports from accredited institutions. Basically, the sampling, test methods, and sufficient information about the test and test conditions need to be included in test reports from the institutions so that they can be rigorously scrutinized by an independent verification Organization.

6.1 Vendors with Certification under other Jurisdictions

The Canadian ETV Program system of verification can be used beneficially by vendors who already have verification in another jurisdiction. This would provide Canadian verification for the Canadian market. In addition, the vendor could use the Canadian ETV Program verification within North America and Europe, or wherever the Canadian ETV Program and Environment Canada are negotiating or discussing harmonization procedures, or have already established harmonization. Several situations may present themselves, including:

16. The technology has third party verification of a performance claim on the basis of a nationally recognized program in another nation or state. In addition there is a documented test program with credible data and the full information is available. In this case, there would be an opportunity for a “fast track” to verification, still employing a Verification Organization and generally employing the process of the Canadian ETV Program General Verification Protocol
17. The technology has a certificate of performance and/or a full report issued by an internationally recognized test laboratory or other institution. In this case the VO and the Canadian ETV Program can verify that the documentation relating to this achievement has been inspected by them, and can include a statement naming the certificate and the issuing institution, and include it as part of the list of Performance Claims.

6.2 Technology That Has Been Recognized For Conformance To Science-Based Regulations

There could be a situation where the technology, or a sample of its product, has been submitted to a government-backed regulatory body for testing for conformance to a specific regulation. This could be, for example, a health-based regulation on purity of a material for potential use as a food or drug; or a hazardous material e.g. a pesticide, and conforms to regulatory standards within a particular national regulatory framework. In this situation, the VO and the Canadian ETV Program can verify that the documentation relating to conformance to such a regulation has been reviewed, and a statement naming the exact regulatory standard can be included as part of the list of Performance Claims.

The utility of this is that the vendor does not need to present a proof of conformance and the

associated documentation in a discussion with a potential customer, and would only present the Canadian ETV Program testament that the documents had been reviewed and that the regulatory conformance is accurately represented in the Performance Claim.

6.3 Participants in a Performance Benchmarking Program

A performance benchmarking program may be initiated where there is a common requirement to examine performance, e.g. for fleet managers wishing to examine various products that would improve fuel consumption and emissions on existing vehicles. Such a performance benchmarking program may be extensive and may be international in scope, e.g. Canada and the United States. In this case the parties involved would agree on performance characteristics to be tested, and would agree on the methodology of testing. A Performance Benchmarking Test Protocol would be developed and all candidates for acceptance under the category would be subject to the same testing. In addition to a technology specific test, a technology specific verification protocol would be developed, for greatest efficiency in completion of the benchmarking. The principles expressed and detailed in the Canadian ETV Program's General Verification Protocol would be followed. This would, in addition to demonstration of compliance with a recognized Government of Canada program, enable the Benchmark Program to more rapidly come to their definition of both the Test Protocol and Verification Protocol, and would thus be both economical and efficient.

6.4 Proof-of-Principle Programs for Pre-Commercial Technology

Technical and scientific inventions go through various stages before full commercialization is possible. However, the resources to continue with development of the technology would, in many cases, be available, only if the technology could provide Proof-of-Principle. The standard Canadian ETV Program Verification protocol requires that the technology be commercially ready, and assumes that the test procedure is performed on the technology or process that will be offered commercially.

When Proof-of-Principle is required, the principles of testing and verification outlined in the General Verification Protocol are equally applicable. However, the specifics of each situation must be considered individually. In addition to the proponent (the developer of the technology, but not the vendor), an appropriate project group must be designated that includes:
The proponent

- An expert independent test agent, who will be involved with the proof of principle test program (or a named laboratory, institution or test agency)
- The Canadian ETV Program
- The Verification Organization
- (a funding program in which the project is supported – optional)

A technology specific test program is to be developed, and approved by designated authorities in the project group. A technology specific verification protocol may also be required. Then, after completion of the experimental test program, a complete documentation package is assembled and the verification review proceeds in the manner herein described.

If the technology or process then proceeds to full commercialization, it may be possible to "fast-track" the verification. In this case, the differences between the pre-commercial testing and the equivalent commercial testing would be identified. This would be followed by a test program that supplied the information that was not otherwise available, and the verification of performance of

the commercialized technology would proceed.

6.5 Technology With Test Data That Is Not Categorized Under the Canadian ETV Program SAW (Statistical Analysis Worksheet)

Not all technologies can be tested in such a way that the data produced will be categorized for examination using the SAWs given in this protocol. In some of these cases, another SAW must be created, or perhaps a different mathematical analysis strategy is required, and the VO and the Canadian ETV Program, aided by an expert statistical consultant will recommend a viable verification analysis procedure.

6.6 Harmonization With Programs Of Verification In Other Jurisdictions

ETV harmonization is a multi-level cooperative initiative to help bring credible technologies that benefit the environment to the forefront by working with ETV partners worldwide. Harmonization can improve access to information on market opportunities while also facilitating a greater understanding of the needs of technology users.

The Canadian ETV Program is also augmenting efforts to support the international harmonization of assessment protocols and test methods, building on the established ETV Generic Test Protocol and other related decision-support tools. The principal elements of this strategy are:

- sharing of protocols and test methods
- mutual recognition or accreditation of verification organizations, and
- country-to-country reciprocity, where practical.

6.7 Notes on related verifications [OPTIONAL]

The VO is advised to review the list of current USEPA ETV verified technologies to identify in the verification report which technologies of a similar nature have been USEPA ETV verified: http://www.epa.gov/etv/verifications/verification-index.html

Related verifications that would be of use to the reader of the Canadian ETV Program Verification Report would be useful to include. Such topics could include similar findings, similar test methods for example.

7.0 The Verification Report

7.1 Establishment of the Audit Trail

7.1.1 Summary of key supporting documents

As a summary of some of the most important paper documents that the VO needs to possess, refer to Table 8

Table 8 Key documents

KEY DOCUMENTS	Present	Absent
Raw data sheets and summary data	☐	☐
Signature pages	☐	☐
Signed Formal Application	☐	☐
Declaration Regarding Codes & Standards	☐	☐
Patent(s)	☐	☐
Sample security: e.g. chain of custody sheets for each sample *	☐	☐
Operation and maintenance manual	☐	☐
Field notebooks	☐	☐
Certificate of accreditation of laboratories	☐	☐

* These items may or may not be available for the Verification Organization but are useful in determining reasons for data discrepancies etc. Where applicable and depending on the nature of the verification test program the VO should request to see these asterisked items.

7.2 Example Report Format for Environmental Technology Verification Reporting

The methodology outlined in this protocol enables the reviewer, the VO, to perform a structured and systematic examination of the field test program and its results. A series of checklists is used, so that many items of review are covered efficiently. The checklists, Tables 1 to 8, are supplied electronically. For the Verification Report, the VO copies the completed checklists

directly into the report. Within each individual question, the VO may add a short explanation for the answer, if needed. More extensive text relating to the topics covered in a particular checklist/Table would be placed adjacent to the Table.

[Within each checklist, there is indication whether the requirement must be fulfilled (mandatory) or is treated as useful or desirable information (optional).]

The following Table of Contents is an example verification report template that may be used to prepare the final verification report. The VO may propose to use other than the following standard format, if prior approval of the Canadian ETV Program is obtained before drafting the report.

Executive Summary

Table of Contents

Introduction
Report format
Background
Objectives
Scope
Legal Notices

Review of Application
- Introduction
- Applicant Organization
- Review of Application
- Application Review checklist comments

Review of Technology
- Technology Review criteria
- Technology Review checklist comments

Review of test plan, test execution and data
- Review of test plan and execution of test plan
- Data validity checklist
- Data validity checklist comments
- Remote sensing data (e.g. telemetry)
- Remote sensing data comments
- Other verification topics
- Other verification topics comments
- Data analysis checklists
- Data analysis checklists comments
- Data interpretation checklist

Data interpretation checklist comments

Statistical Analysis of the performance claim(s)

- Performance claim
- Performance claim(s) verification

Establishment of the audit trail
Audit trail comments

Conclusion

APPENDICES
Appendix A Statistical Analysis Worksheets
Appendix B Statistical tables
Appendix C Declaration regards codes and standards
Appendix D Append Critical material as necessary

FIGURES
LIST FIGURES AS NECESSARY

TABLES
LIST TABLES AS NECESSARY INCLUDING THE FOLLOWING:

Table 1 Application Review Checklist
Table 2 Technology Review Criteria Checklist
Table 3 Verification Study Design Assessment Criteria Checklist
Table 4 Data Validity Checklist
Table 5 Data Analysis Checklist
Table 6 Data Interpretation Criteria
Table 7 Technology/Input Criteria
Table 8 Summary of Data Sets Submitted
Table 9 Results of Statistical Analyses
Table 10 Summary of Claim Evaluations

Environmental Technology Verification

General Verification Protocol

Appendix A

Statistical Analysis Worksheets
(for Performance Claim Verification)

February 2007
(Revised May 2013)

Environment Canada Environnement Canada

TABLE OF CONTENTS

1.0 Introduction to the Statistical Analysis Worksheets (SAWs) 2
2.0 Standards and Conventions of the ETV SAW Package 3
3.0 Background Information and Examples 4
3.1 Introduction to SAWs 4
3.2 Introduction to Distributions 4
3.3 The Normal Distribution 5
3.4 The t-distribution 6
4.0 Experimental Design 7
4.1 Populations and Samples 7
4.2 Statistical power 7
4.2.1 Sample size estimation – An Example 7
4.2.2 Example of estimating sample size for One Sample T-Test 10
4.3 Replication and Pseudoreplication 12
4.4 Independence of Observations and Data Sets 12
5.0 Introduction to Statistical Inference 14
5.1 Statistical Significance 14
5.2 Hypothesis Testing 14
5.3 The Confidence Interval 15
5.4 A General Paradigm for the Statistical Assessment of Performance Claim 15
5.5 Two-Sided versus One Sided Tests 16
5.6 Using the Statistical Tables 17
6.0 Statistical Analysis Work Sheet (SAW) 18
6.1 SAW # 1 Assessing Normality of Data 18
6.2 SAW # 2 Calculation of a 95% Confidence Interval for a Mean 21
6.3 SAW # 3 Testing Equality of Two Variances 23
6.4 SAW # 4 Testing Percentage Reduction 24
6.5 SAW # 5 Testing Mean is Equal to a Specified Value 26
6.6 SAW # 6 Testing Equality of Two Means when Sample Variances are Assumed Equal 28
6.7 SAW # 7 Testing Equality of Two Means when Sample Variances are Assumed Unequal 31
6.8 SAW # 8 Testing Median is Equal to a Specified Value 34
6.9 SAW # 9Testing Equality of Two Medians 36
6.10 SAW # 10 Testing For Mean Difference in Paired Observations 38

Introduction to the Statistical Analysis Worksheets (SAWs)

The SAWs provide a limited introduction to the host of statistical tools that may be used to test a performance claim. The SAWs used in this General Verification Protocol (GVP) were chosen to provide analytical methods for the most common types of data sets generated by verification test programs. They are suitable for use by the non-statistician, provided test assumptions are verified and the concepts emphasised in this SAW are understood and used when data interpretations are made. Other numerical and analytical methods may be added, at the discretion of the Canadian ETV Program, or as new statistical methodologies are developed for the GVP. It is the responsibility of the Verification Entity to obtain approval from the Canadian ETV Program prior to using alternative procedures.

It should be emphasized that a verification claim could be tested in several ways. The method of choice is the simplest, and the most defensible. The term "defensibility" is used in a legal sense. Errors or oversights in experimental design, or data analysis could render a verification void.

The phrase "experimental design" for evaluating performance encompasses the concepts of

- statistical power[1],
- experimental unit[2],
- acknowledgement of interferring factors,
- the method that will be used to interpret the data generated, and
- a thorough knowledge of the technology undergoing verification testing.

The phrase "data analysis" encompasses the recognition of the experimental unit and confounding factors, choice of analytical method, assessment of assumptions made and correct interpretation. Although the SAWs incorporate tests of assumptions and the rationale for use, it is still possible to perform a technically correct data analysis and, at the same time, have an unsubstantiated performance claim (because other mandatory conditions have not been fulfilled).

A statistical glossary is provided that contains commonly used terms and should be used to clarify concepts. The following concepts constitute the minimum level of understanding required to apply the SAWs:

- An understanding of the relationships between an experimental unit, a variable, an observation and a replicate.
- The difference between a null and alternative hypothesis.
- Independence of observations and experiments.

[1] Refer to section 4.2
[2] Refer to section 4.3

1.0 Standards and Conventions of the ETV SAW Package

For the Canadian ETV Program, and other ETV programs, the industry standard is 95% confidence.
An α (alpha) value of 0.05 is required for significance tests or confidence intervals. This alpha value relates to "95% confidence intervals or 5% significance" because (1-alpha), on a percentage basis is 95%.

A small sample size is described as less than or equal to 30. Users with access to statistical software may wish to increase this limit to 60.

The number of digits to report when performing calculations is 1 more than that obtained for an observation. When performing calculations, this rounding is only performed at the final step. This procedure is shown in the case studies, Appendices E and F.

References

1) A useful introductory textbook that provides statistical fundamentals and general information is:

McClave, T., and Sincich T. Statistics, 9th edition, Prentice-Hall Inc. New Jersey ISBN 0 – 13-065598 – 8 (2003)

Steel, R. G. and J. H. Torrie. Principles and Procedures of Statistics. McGraw-Hill Book Company, New York. (1980)

2) Design of experiments and additional reference materials

Box, Hunter and Hunter. Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building. John Wiley and Sons, New York. (1978)

Snedecor, G. W. and Cochran. W. G.. Statistical Methods. Iowa State University Press, Ames Iowa. (1980).

Limitations

The SAWs are used as a tool to aid the VE in evaluating the data supplied by the applicant or a testing agency and to determine whether the data verify the environmental technology performance claim(s) made by the applicant. Best professional judgement should be employed by the VE at all stages of verification program including the statistical evaluation of the data. Other methods or procedures must be approved by the Canadian ETV Program prior to use in a verification by the VE.

2.0 Background Information and Examples

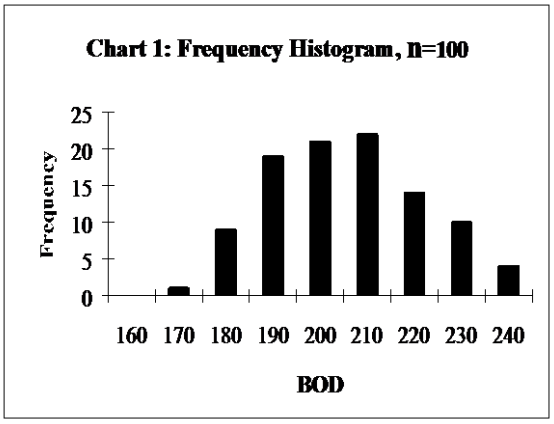
2.1 Introduction to SAWs

This section explains some of the statistical precepts underlying the SAWs and provides examples of statistical calculations. While mastery of the material in this section is not required to complete the SAWs, an increased awareness of fundamental issues will provide confidence to the user and aid in improving the design of verifications studies.

Note, the statistical analysis package used in **Appendix A – SAWS,** and illustrated by example in **Appendix H – Examples of Statistical Calculations** is Microsoft Excel worksheet based. However, packages such as Minitab®[3] Statistical Software can also be used for the same statistical analysis of data that is defined by the SAWs for verification of performance claims.

2.2 Introduction to Distributions

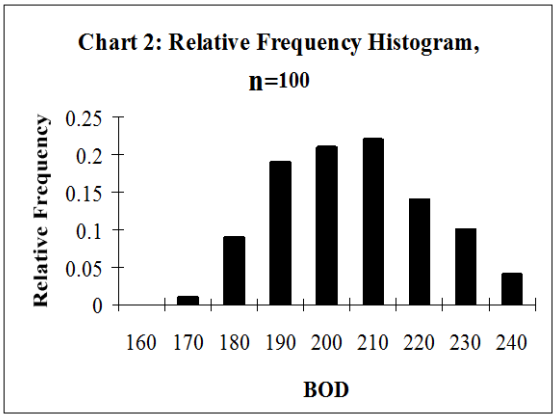
Many of the processes being verified by the ETV program generate measurements or data that are continuous. ***Continuous data*** *can take on an infinite number of values such as fractions or integers and may include negative numbers.* Some examples of continuous data are height, mass, chemical concentration, etc. These variables are measured, not counted.



If we were to sample a process that generated continuous data, we would often find that many of the values were similar in size while some of the values would be quite different from the majority of the values. *If we subtract the smallest value from the largest value we have estimated the* ***range*** *of values the variable may take*. For example a chemical measurement such as BOD may range from 160 to 240 mg/l with most of the values clustered around 200 mg/l (total number of observations, n=100). The range of this data set is 80 mg/l. If we break up this range into 8 bins or classes of 10 mg/l we can count the number of observations that fall into each of the bins or classes. We expect that most of the values will fall into the class 190-200 mg/l and 200-210 mg/l. A much smaller number of values will fall into the class 160-170 mg/l or 230-240 mg/l. A bar plot of these values is known as a **frequency histogram** (Chart 1). *It describes how many observations fall into each class or bin.*

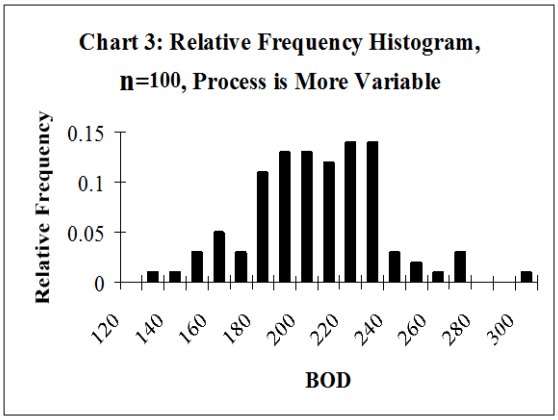
[3] Minitab® Statistical Software, Minitab Inc. State College PA. URL: www.minitab.com

If we divide the number of values falling into each bin by the total number of values, we obtain the proportion of values falling into each bin known as relative frequency (Chart 2). It might also be called a probability of falling into a bin. We note that the relative frequency histogram is the same shape as the previous histogram, only the axis has changed.



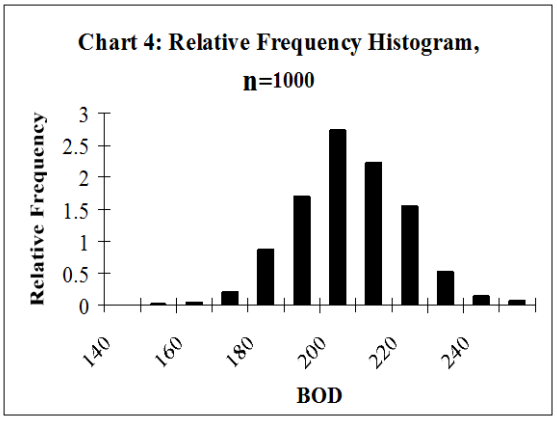
The histogram tells us how likely a value is to occur. *The most likely value to encounter when randomly sampling the BOD data set seems to be the* ***mean*** *or* ***average*** *BOD. The mean* (defined in Appendix D) *is usually designated by the symbol* ***μ****.*

Now consider a process that induces variability in the BOD data set. We see that the distribution is more spread out (Chart 3). The mean is still the most likely value to be measured by random sampling. However, compared to the more peaked curve (Chart 2), the probability of detecting the mean is less. Thus, the more variable the process, the less confidence we have in our measurement (for the same n). *This measurement of the spread around the mean is known as the* ***variance*** *and is designated as* $\boldsymbol{\sigma^2}$. *The square root of the variance is often encountered in statistics and has it's own name, the* ***standard deviation****. The standard deviation is designated as* $\boldsymbol{\sigma}$.



2.3 The Normal Distribution

The normal distribution is one of the most commonly encountered in statistics. Many continuous measurements, i.e. those that can take on an infinite number of values such as a height, mass, chemical concentration, etc. are described by this distribution. We can draw a frequency histogram for this distribution as before. The following histogram (Chart 4) is generated using 1000 points, but has the same mean and standard deviation as the frequency histogram in Chart 2. *The* ***normal distribution*** *is "bell-shaped" and roughly symmetric about the mean value.* While very large or small values are unlikely, there is a possibility that such values could occur.

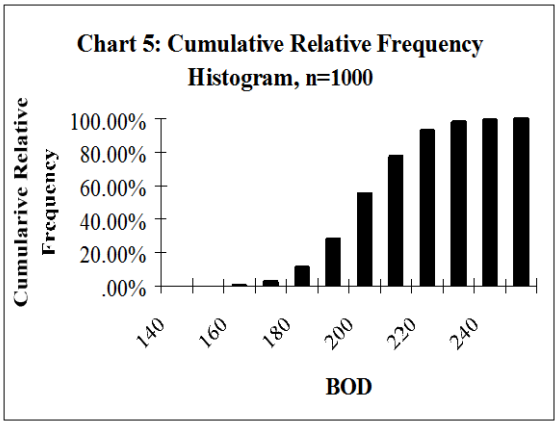


If we have a sample and wish to determine how likely it is that the mean value is less than a given or hypothesised value, we can use a frequency histogram as above. For example if we design an effluent process that claims to produce an output BOD of less than 150 mg/l, and the experiment generated the data above, we would say that this claim is unlikely to be

true since most of the measured data fall above the value of 150 mg/l. Of course we could repeat the experiment generating the data, but it is unlikely that the frequency histogram will shift to the extent that a mean BOD value of 150 mg/l is likely.

Now, if we consider the claim that the effluent process produces an output BOD of less than 240 mg/l, we would be much more likely to support this claim. We would do so, since from our sample, many of the values, in fact most of the values fall below 240 mg/l.

If we were to add up all the probabilities reading from left to right, and re-plot the frequency histogram we would obtain a **cumulative relative frequency histogram**.



We could use Chart 5 to determine that the probability of a value falling below 240 is roughly 99.4%. Thus if we made the performance claim that the process produced an output BOD of 240 mg/l or less, we would be 99.4 % certain that we were correct. Cumulative relative frequency histograms are very useful for making objective decisions about a data set. However generating such a histogram is time consuming and for small data sets not very informative. If we can assume that our measurements arise from a normal distribution, then we can compare the values obtained from our experiment with those that would be expected from a normal distribution with the same mean and variance.

The generation of values from a normal distribution requires some tedious calculations. Therefore statisticians have generated tables of these values. As the number of tables required to describe the frequency distribution of all possible combinations of means and variances is infinite, a standardized table has been produced.

The standardized table assumes that the normal random variable designated as Z, has a mean of zero and a standard deviation of 1. The tests used in the SAWs convert the measured values, so that the converted values have a mean of zero and a standard deviation of 1. This allows a single table to be used for all variables that have, or are assumed to follow a normal distribution. If we designate our converted random variable as z (note lower case) then the standard normal random table provides the probability that our standardized normal random variable z, is less than or equal to some value Z, that follows the standard normal distribution. This is exactly the same process used to determine the probability that an output BOD value would be less than 99.4%. This statement may be made more succinctly as P(z = Z).

2.4 The t-distribution

Although many measurement variables follow a normal distribution, it was observed in the early part of the 20th century that when sample sizes were small, the probability of obtaining very large or very small values (i.e. extreme values) was underestimated when a normal distribution was assumed. The ETV program requires that the small sample approximation to the normal distribution, known as the t-distribution, be substituted for the normal distribution when the sample size is less than 30 in accordance with generally accepted statistical usage.

3.0 Experimental Design

3.1 Populations and Samples

The performance claim requires that a data set be generated to test the claim. This data set comprises only a small fraction of all the possible outcomes of the technology being evaluated. *This fraction of all possible outcomes is known as a* ***sample***, whereas *the set of all possible outcomes is known as the* ***population***. For practical reasons (of which cost is one), we almost always make inferences regarding the population, based upon a sample rather than sampling the entire population.

3.2 Statistical power

In the ETV program, statistical proof at the 95% confidence level is accepted as the necessary condition for verification of performance, where the performance is shown by a data set representing the results of operating the technology under some specified operating conditions. By this means, we know that the probability that the performance result that has been verified is due to chance is very low, less than five percent.

The probability that we will be able to show the performance target, with the verification test that has been performed, must be high, and this is known as the statistical power of the test. To examine the power of the test, we consider four inter-related parameters – the statistical significance, the standard deviation, the sample size and the power of the test. There is a relationship between the four parameters that allows any one of them to be calculated if the other three are known. In other words, we want to avoid the situation in which the performance claim could potentially have been verified if only more samples had been acquired. As we will understand intuitively, if there is great variability in the data, and/or the data is very close to the regulatory limit which is the numerical value stated in the hypotheses, then a performance claim verification may "falsely" be declared to be unsuccessful.

3.2.1 Sample size estimation – An Example

The following example of sample size calculations is based on a technology treating potable water to produce chromium (Cr) effluent that should meet a regulatory standard of 50µg/L

The choice of sample size (i.e., number of measurements of the key parameter – Cr concentration) depends upon

1) the initially unknown variability among observations, and
2) the degree of certainty required, when testing hypotheses.

The following two paragraphs describe the concepts involved in making *a priori* sample size decisions.

A type I error is the probability of rejecting the null hypothesis when it is true. This probability is commonly designated as "alpha" (α). In the context of this experiment, there would be a type I error if one concluded that a treatment produced a mean or median effluent chromium concentration <50 µg/L, when it did not. This error would cause unnecessary

additional testing. The probability of this error occurring is called the level of significance. The ETV program has set this level to 5%. The probability of not making this error is called confidence (1-alpha). The program is designed for 95% confidence.

A type II error occurs if we accept the null hypothesis when it is false. The probability of this type of mistake occurring is commonly designated as "beta" (β). In the context of this experiment, one would conclude that the mean or median effluent Cr concentration is greater than or equal to 50 µg/L when the mean or median effluent Cr concentration is really <50 µg/L. Therefore a potentially promising technology would appear to be less effective. ***The probability of not making this error is called power (1-beta).*** The example here is for 90% power in hypothesis testing. However, the generally accepted value is 80% statistical power.

The stated α and β values are widely used by scientists as acceptably low probabilities of type I and type II errors. Somewhat lower or higher values are sometimes chosen, based on consideration of the consequences of these two types of errors in decision making.

If the **null hypothesis is H_o:** Mean effluent Cr concentration ≥ 50 µg/L and the **alternate hypothesis is H_a:** Mean effluent Cr concentration <50 µg/L, then, the equation for sample size estimate is:

$$n = (S(Z_\alpha + Z_\beta)/d)^2$$

where:

- S = standard deviation (µg/L)
- d = difference between mean and 50 µg/L
- Z_α = normal deviate value for tail probability (refer to GVP Appendix B Table B1)
- Z_β = normal deviate value for tail probability (refer to GVP Appendix B Table B1)

When the current estimate of mean and variance are based on a small number of samples (n_c), the estimate of required n should be adjusted up by a factor of $(n_c + 2)/n_c$.

Table.1 shows the calculated sample size requirement for different values of mean and standard deviation, and 10% error probabilities. This table balances the probability that technologies are unnecessarily tested (type I error) with the probability that promising technologies are deemed less effective than they really are (type II error), by choosing reasonably small values for each of these probabilities.

Table 1: Sample Size for 10% Probability of Type I and Type II Errors

Standard Deviation	Observed Mean	Probability (Type I error)	Probability (Type II error)	Required Sample Size
5	48	10%	10%	42
10	48	10%	10%	165
5	45	10%	10%	7
10	45	10%	10%	27
5	40	10%	10%	2
10	40	10%	10%	7
20	40	10%	10%	27

This table shows that the number of samples depends upon:

- The degree of variability (standard deviation) among the observations. When the test commences, this value is unknown.
- The degree of difference between the hypothesized mean (50 µg/L) and the observed mean. This value varied from 48 to 40 in the table above.

To detect a small deviation (such as 5 µg/L) in effluent Cr, concentration from 50 µg/L, in a situation where the observations are variable, 27 observations would be required. To detect a large difference between the effluent Cr concentration and the cutoff of 50 µg/L, if there is little variability in the data set, only 2 observations would be required (Table 1).

Table 2 shows the effect on the sample sizes of relaxing probabilities of type I and type II errors.

Table 2: Sample Size for 40% Probability of Type I and 20% Probability of Type II Errors

Standard Deviation	Observed Mean	Probability (Type I error)	Probability (Type II error)	Required Sample Size
5	48	40%	20%	8
10	48	40%	20%	30
5	45	40%	20%	2
10	45	40%	20%	5
20	45	40%	20%	20
5	40	40%	20%	1
10	40	40%	20%	2
20	40	40%	20%	5

The preceding two tables indicate that the confidence level (1 - α) and the power (1 – β) achieved by the test have to be estimated after the test has been completed, and the data have been obtained. Also, the achieved type I error rate or p-value is obtained after the test has been completed.

Table 3 shows the sample sizes required to determine a difference of 5, 15 and 25 µg/L from a hypothesised value of 50 µg/L with standard deviations in increments of 1. The probabilities of type I and type II errors are both set at 10%.

Table 3: Sample Size for 10% Probability of Type 1 and II Errors for Observed Means of 45 µg/L, 35 µg/L and 25 µg/L

	Observed mean = 45 µg/L	Observed mean = 35 µg/L	Observed mean = 25 µg/L
Standard Deviation	Required Sample Sizes		
1	1	1	1
2	2	1	1
3	3	1	1
4	5	1	1
5	7	1	1
6	10	2	1
7	13	2	1
8	17	2	1
9	22	3	1
10	27	3	2
11	32	4	2
12	38	5	2
13	45	5	2
14	52	6	3
15	60	7	3
16	68	8	3
17	76	9	4
18	86	10	4
19	95	11	4
20	106	12	5

A probability value, or p-value in common usage, is the probability of observing the experimental mean, assuming that the null hypothesis is true. A test is formally rejected if this p-value is smaller than the level of significance decided upon before the experiment is conducted. For example a p-value of 0.023 (2.3%) would lead to rejecting a null hypothesis when the probability of a type I error was set to 0.05 (5%).

3.2.2 Example of estimating sample size for One Sample T-Test

Purpose: This analysis is used to estimate the sample size required to achieve pre-specified type I and type II error rates when comparing the mean to a pre-specified value using a one-sample t-test.

Alternative Hypothesis:
The alternative hypothesis is that the mean effluent Cr concentration is < 0.05 mg/L.

User Notes: Type I error = 5%
Type II error = 10%

Only those values required for calculations are presented below

Table 4: Observations

	Effluent Cr (mg/L)	Squared Effluent Cr
	0.0606	0.00367
	0.05278	0.00279
	0.06495	0.00422
	0.07321	0.00536
	0.07259	0.00527
	0.07687	0.00591
	0.04553	0.00207
	0.06113	0.00374
Sum	0.50765	0.03302
N	8	
Mean($\bar{x}$)	0.06346	
Sample variance (σ^2)	0.01075	
Z_α	1.645	
Z_β	1.285	
μ	0.05	

Steps to follow using Excel:

1. Square all observations.
2. Sum the observations.
3. Sum the squared observations
4. Estimate the mean of the observations
5. Estimate the sample variance (sigma^2)

Using Table 1 in GVP-Appendix B:

6. Find Z_α and Z_β values where α = 5% and β=10%
7. Calculate sample size using the following formula

$$n = \left(\sigma^2 \times (Z_\alpha + Z_\beta) \Big/ |\mu - \bar{x}| \right)^2$$

$$n = \left(0.01075 \times (1.285 + 1.645) \Big/ |0.05 - 0.06346| \right)^2 = 5.47$$

When the current estimate of mean and variance are based on a small sample size (n_c), the estimate of n should be adjusted up by a factor of (n_c + 2)/n_c. This factor is equal to 1.25 for n_c = 8 and for this example, the final value is *n* = 6.8

Interpretation: We require 7 samples to ensure a type I error not greater than 5% and a type II error not greater than 10%.

There are 8 observations (samples) in the experiment. The null hypothesis, that the mean is greater than or equal to 0.05 mg/L, is therefore accepted. The probability of error is not greater than 10%.

3.3 Replication and Pseudoreplication

*An **experimental unit** is the smallest unit to which an experimental treatment may be applied.* For example if we are testing a new type of technical equipment, then a second identical equipment unit tested in the same way constitutes a replicate. Taking another type of example, if we apply an environmentally friendly defoliant to a field, then a field with the same application of defoliant constitutes a replicate. If we wish to repeat the experiment to obtain a replicate we could not (easily) apply the defoliant to the same field. We would need to apply the defoliant to a different field. Measurements of parameters such as percent weed cover for plots within the field **are sub samples or pseudoreplicates**. These measurements are not replicates.

In order to make valid statistical inferences, we must measure the variability among replicates. In the previous example, the variability among plots measures the variability within the field. This estimate of variance is likely smaller than the variability among fields. The inadvertent use of a variance estimated from sub samples or pseudoreplicates will bias the variance estimate downward and artificially increase the power of the verification experiment. Examples of calculations are found in Appendix H, and the concept of statistical power is outlined in section 4.2.

The practical implication is that we will be able to detect smaller differences between two different treatments, for example weed cover with and without a defoliant. If the endorsement of our technology depends upon finding a difference between a reference or standard condition and the new technology, we will find differences more often than we should. **For this example, the technology may not be as successful as was claimed when applied to the intended population.**

Some SAWs are designed to use data from replication testing, and the benefits are:

- As we increase the number of replicates we increase our coverage of the sample population. This allows us to make inferences of broader scope. When marketing a technology, generally, the less restrictive the claim is, the better.

- As we better define the variability of the experiment we are able to make more precise claims. If all other parameters are equal, a technology with a small variance (or in more pragmatic terms, a more proven track record) will better meet expectations.

3.4 Independence of Observations and Data Sets

The concepts of independence of observations and data sets are closely related to the definition of a replicate. *Observations are generally **independent** if they are obtained from different experimental units. In other cases, observations obtained from a process may also be independent if they are sampled far enough apart in space or time*.

For example, consider measuring a parameter in a wastewater stream. If we take samples separated by a very short period of time, we would expect that measurements would be similar to one another. As the time between sampling events increases, the measurements become less and less dependent upon the previous measurement. This type of dependency is known as ***correlation*** or more specifically autocorrelation (correlation in time). Correlation may occur in time as described above, or spatially. Data of this type should not be analysed using the ETV SAWs. The VE should understand the process sufficiently to determine whether correlation between measurements is likely. If measurements are correlated, expert advice should be obtained.

Data sets may also be dependent upon one another. As an example consider the application of a treatment to a split sample. If we applied an environmental technology to both halves of the split sample, we would expect the results to be more similar than the results from two different samples from the technology.

Generally we apply a technology to one half of a split sample and not to another half, as we wish to compare the difference in effects between each of the split samples. The two observations or measurements arising from a split sample are known as paired observations. This type of experiment is known as a paired experiment. The analysis of paired data is described in SAW #10.

4.0 Introduction to Statistical Inference

4.1 Statistical Significance

The concepts of statistical significance can be illustrated using the BOD example. We saw that given the performance claim that an effluent produces a BOD less than 240 mg/l we would be 99.4 % certain that we correctly verified the claim. Upon accepting the claim, we would also be 0.6% certain that we had incorrectly verified the claim. This is true because 0.6% of our representative sample is greater than 240 mg/l. This type of error, when testing a statistical hypothesis, is known as a **Type I error** and is designated as **α**. (**1-α)** is also described as the level of significance or level of statistical significance. Traditionally, this value has been set to 95 %. An **α** value of 0.05 is required by the ETV program when testing a performance claim. Where appropriate, this default value has been inserted into the SAWs.

4.2 Hypothesis Testing

When sampling from a population, it is not possible to prove that a performance claim or hypothesis is true. We can only disprove hypothesis and accept an alternative hypothesis or performance claim. Therefore we must pose a performance claim in such a way that the process of disproving it, verifies the claim. *The hypothesis being disproved is referred to as a* ***null hypothesis*** and *the hypothesis being accepted, after rejecting the null hypothesis is the* ***alternative hypothesis***. The alternative hypothesis expresses the essence of the performance claim. These two hypotheses are designated as $\mathbf{H_o}$ and $\mathbf{H_a}$, respectively. An example will serve to clarify these concepts.

In the previous example, the performance claim was that the process generating the effluent would result in a BOD of 240 mg/l or less. Our null hypothesis H_o is:

The process produces an effluent with a mean BOD equal to 240 mg/l,

or more succinctly as:

H_o*:* μ_{BOD}*=240 mg/l.*

The alternative hypothesis (what we are interested in accepting, not proving) may be phrased as:

The process produces an effluent with a mean BOD less than 240 mg/l,

or more succinctly as:

H_a*:* $\mu_{BOD} < 240$ *mg/l.*

Hypothesis testing leads to a ***decision*** to either **reject** or **not reject** the null hypothesis. If we can reject the null hypothesis, we can accept the alternative hypothesis and therefore verify the performance claim. The hypothesis testing is completed by stating a ***conclusion***. Depending on the decision made regarding the null hypothesis, the conclusion is formulated in the following manner: "There **is** sufficient evidence at the 95% level of significance to reject the null hypothesis. We conclude that …**the meaning of the alternative hypothesis**".

OR:

"There **is insufficient** evidence at the 95% level of significance to reject the null hypothesis. We conclude that … **the meaning of the null hypothesis**".

4.3 The Confidence Interval

At times we may not wish to test a hypothesis. Instead we want to ascertain the degree of confidence regarding some performance characteristic of an environmental technology. In section 3.3, we showed how a cumulative frequency histogram could be used to determine that the probability of a BOD value falling below 240 is roughly 99.4%.

In a similar way we could ask the question: For what range of values would I be 95% certain that my true process measurement would fall?
Notice that the range of values falls around the true process characteristic not the sample measurement. We do not know what the value of the true process characteristic is. We conduct a verification experiment to estimate this value. This estimate is only based upon a sample of all possible outcomes. Obviously, as we increase the number of samples so that we are sampling a larger proportion of all possible outcomes we are more and more certain of our estimate. If we sample the entire population, there is no uncertainty about our measurement and hence no need for a confidence interval. The difference between a sample and a population is described in section 4.1.

By assuming that our sample estimate arises from a given statistical distribution we can determine the range of values for which we could say: *"I am 95% certain that the unknown process measurement lies within these two values".*

This is known as a **confidence interval**. If we wish to be very certain about a measurement we can increase the level of significance and consequently, the size of the confidence interval increases. A confidence interval may also be very wide, if the measurements used to estimate the statistic of interest are extremely variable. Thus a highly variable sample will produce a larger confidence interval than a less variable sample. The variability of a sample may be reduced by increasing the sample size or controlling sources of variability through the experimental design and/or analysis.

The validity of the confidence interval depends upon assuming the correct statistical distribution for the measurement around which the confidence interval is being constructed. The ETV program requires that distributional assumptions be tested prior to constructing a confidence interval.

4.4 A General Paradigm for the Statistical Assessment of Performance Claim

The following text summarises the general outline of a performance claim using terminology defined in the previous sections. An understanding of this terminology is helpful in interpreting the SAWs or when soliciting advice from a statistician.

To verify a performance claim we generate a data set that meets the criteria set out in Tables 3 and 4 of the GVP. *The process of generating a data set is referred to as an experiment or a* ***verification test program*** *in the context of this documentation.* The experiment generates a data set that is only a fraction of all possible outcomes. This data set is a sample. The data set measures the attribute of some variable such as BOD, in the previous example. Each discrete measurement is known as an observation.

Using the performance claim we formulate a null hypothesis, which we are attempting to disprove. We state the hypothesis we wish to accept as the alternative hypothesis. By knowing or assuming the frequency distribution of the variable being measured, we can determine how probable our performance claim is.

The SAWs provide methods for testing the following types of hypotheses:

- A mean is equal to, less than or greater than a specified value.
- A median is equal to, less than or greater than a specified value.
- Two variances are equal to one another, or one variance is less than or greater than another variance.
- Two means are equal to one another, or one mean is less than or greater than another mean.
- Two paired means are equal to one another, or one of the paired means is less than or greater than the other paired mean.

4.5 Two-Sided versus One Sided Tests

When we compare a sample mean, for example, to a specified value we may wish to test that it is different than the specified value. In this case we would say our sample mean is different from the specified value if the sample value is much smaller or much larger than the specified value.

Since we determine the probability of a certain value occurring from the ends or tails of a statistical distribution, and we are considering both large and small values, we call this ***a two-tailed test***. In this case the hypotheses would be formulated as follows:

Null Hypothesis: Sample Mean=hypothesized value.

Alternative Hypothesis: Sample Mean≠hypothesised value.

If we were interested in determining whether our sample mean was larger than a specified value, we would test the null hypothesis:

Sample mean = hypothesised value.

Our alternative hypothesis would be:

Sample mean > hypothesised value.

We would only reject the null hypothesis if our sample mean were greater than some predetermined (hypothesized) value. The size of this value is determined by our required

degree of confidence in our assertion. For the ETV program this value is 95%. Thus we obtain the value or quantile from the specified distribution for which only 5% of the values are likely to be greater. If our sample value is larger than this cut-off value or critical value, we are reasonably certain that our sample mean, really is larger than the hypothesised value. *Since we used only one tail or end of the probability distribution, this is known as a* ***one-tailed test****.*

We could also test the converse, that the sample mean < the hypothesized value. In this case we would choose the quantile from the specified distribution for which only 5% of the values are likely to be lower.

As in the case of constructing a confidence interval, the assumed distribution must be correct or the test of hypothesis will be invalid. The ETV program requires that distributional assumptions be tested prior to testing a hypothesis whenever a distribution is assumed.

4.6 Using the Statistical Tables

Statistical tables are provided by the Canadian ETV Program in Appendix B. Examples on specific usage may be obtained from the electronic files containing examples for each SAW.

5.0 Statistical Analysis Work Sheet (SAW)

5.1 SAW # 1 Assessing Normality of Data

This procedure is used to determine if the data variable is normally distributed or log-normally distributed. This is important as the assumption of normality is often invoked in subsequent calculations.

Assumptions:
The x_i observations constituting the data set are independent[4].

Data Description	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database

Determining Potential Normality of Distribution	
Data points may be any real number and the range of possible values is infinite. This is often not the case for a measured value such as a concentration, which cannot be negative. In this case it is sufficient that the majority (95%) of the points lie within 3 standard deviations[5] of the mean of the measured points.	o True
The data points are not proportions[6], rates or frequencies.	o True
The data points are not counts.	o True
Is the mean approximately the same as the median? median = mean =	o True
Based on guidelines above, the sample is potentially normally distributed.	o True o False
If the sampling distribution is potentially normal, and there are more than 10 data points, prepare a normal probability plot of the raw data	

Preparation of Normal Probability Plot (if n > 10)	
Order the data (x_i) from smallest to largest. Subsequent calculations use the ordered data.	
Sample size:	n:
Calculate "Blom" coefficients. $p_i = \frac{i - 3/8}{n + 1/4}$, for i = 1 … n.	p_i: unnecessary to present the n coefficients here. Attach a table or spreadsheet.

[4] A non-rigorous definition of independence is in section 4.4.
[5] Standard deviation is defined in Appendix D.
[6] Proportions, rates and frequencies are variously defined. We use these terms to describe a set of numbers that may take on any value between 0 and 1, inclusively

Convert "Blom" coefficients to y_i. $y_i = \sqrt{-\ln(4p_i(1-p_i))}$, for i = 1 … n.	y_i: unnecessary to present the n coefficients here. Attach a table or spreadsheet.
Calculate normal scores. $z_i = sign(p_i - 1/2) \bullet 1.238 \bullet y_i \bullet (1 + 0.0262 y_i)$, for i = 1 … n, where *sign* $(p_i$-$1/2)$= -1, for $(p_i$-$1/2)$<0, *sign*$(p_i$-$1/2)$= +1 for $(p_i$-$1/2)$>0, and *sign*$(p_i$-$1/2)$=0 for $(p_i$-$1/2)$=0.	z_i: unnecessary to present the n coefficients here.
Plot the normal score data against the ordered data.	

Q_1. Do the data appear to fall on a straight line? o Yes o No

If yes, proceed to formal test of normality.
If no and "tails" of distribution fall off the straight-line, log-transform the data and re-plot.

Q_2. Do the log-transformed data appear to fall on a straight line? o Yes o No

If yes, proceed to formal test of normality.
If no, use a test that does not assume normality. For example SAWs #8 and 9.

Test of Normality	
Estimate the Test Statistic	
$SS_{xz} = \sum_{i=1}^{n} x_i z_i - \left[\left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} z_i \right) / n \right]$	SS_{xz}:
$SS_x = \sum_{i=1}^{n} x_i^2 - \left[\left(\sum_{i=1}^{n} x_i \right)^2 / n \right]$	SS_x:
$SS_z = \sum_{i=1}^{n} z_i^2 - \left[\left(\sum_{i=1}^{n} z_i \right)^2 / n \right]$	SS_z:
Estimate Shapiro-Francia W. $W = \frac{SS_{xz}^2}{SS_x SS_z}$	W:
Apply Box-Cox Transformation	
u = ln(n)	u:
v = ln (u)	v:
$\hat{\mu} = -1.2725 + 1.0521(v - u)$	$\hat{\mu}$:
$\hat{\sigma} = 1.0308 - 0.26758(v + 2/u)$	$\hat{\sigma}$:

Transform W to Z′. $Z' = \frac{\ln(1-W) - \hat{\mu}}{\hat{\sigma}}$	Z':
If Z' > 1.645 we reject the null hypothesis that the data are normally distributed at the 95% level of confidence. The data are not normally distributed.	

Q_3. Do the data pass a goodness of fit test[7] for normality? o Yes o No

If answers to questions Q_1 or Q_2 and Q_3 are yes, the raw (or log-transformed) data are normally distributed. The raw or log-transformed data may be used in SAWs assuming normality.

The raw data are Normally Distributed? o Yes o No
The log-transformed data are Normally Distributed? o Yes o No

You can now proceed to the next appropriate SAW.

[7] Recommended test of normality for manual calculations is the Royston modification of the Shapiro-Francia test. Users with access to statistical software are advised to use the Shapiro-Wilks test.

5.2 SAW # 2 Calculation of a 95% Confidence Interval for a Mean

This test is used to determine at a level of 95% confidence that the true but unknown population mean lies within the constructed interval.

Assumptions:

- *The data set is normally distributed.*
- The x_i observations constituting the data set are independent[8].

Data Description and Tests of Assumptions	
Parameter	Units:
Data Location	o attached page
Filename and Location	o electronic database
Based on SAW #1, the data set is normally distributed.	o Yes

Common Calculations	
Estimate of μ	$\bar{x}$:
Total sample size n	n:
Estimate of σ^2 $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$	s^2:
If n ≥ 30	
Obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	$Z_{0.975}$: 1.96
Lower Confidence Limit: LCL = $\bar{x} - Z_{0.975}\, s/\sqrt{n}$	LCL:
Upper Confidence Limit UCL = $\bar{x} + Z_{0.975}\, s/\sqrt{n}$	UCL:
If n <30	
Obtain $t_{0.975,\, n-1}$ from Table B2, Appendix B, GVP.	$t_{0.975,\, n-1}$:
Lower Confidence Limit LCL = $\bar{x} - t_{0.975,\, n-1}\, s/\sqrt{n}$	LCL:

[8] A non-rigorous definition of independence is in section 4.4.

Common Calculations	
Upper Confidence Limit UCL = $\bar{x} + t_{0.975,n-1} \frac{s}{\sqrt{n}}$	UCL:

The 95% confidence interval for the mean $\bar{x}$ is: (LCL, UCL).

5.3 SAW # 3 Testing Equality of Two Variances

H_o: $\sigma_1^2 = \sigma_2^2$

This test is used to determine at a level of 95% confidence, if two variances are equal. The equality of variances is important when pooling data sets. The formulae presented below are applicable when the two data sets are equal or unequal in number.

Assumptions:

- *Both data sets are normally distributed.*
- *The x_i observations constituting the data set are independent[9].*
- *Data sets are independent of one another[10].*

Data Description and Tests of Assumptions	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database
Based on SAW#1, the data sets are normally distributed.	o Yes

Common Calculations	
Estimate of σ_1^2 (Let larger variance correspond to numerator)	s_1^2:
Estimate of σ_2^2	s_2^2:
Degrees of Freedom Data Set 1 = n_1 - 1	$v_{1:}$
Degrees of Freedom Data Set 2 = n_2 - 1	v_2:
Test statistic F = σ_1^2 / σ_2^2	F:
Calculations Case A - H_a: $\sigma^2_1 \neq \sigma^2_2$	
Obtain $F_{0.975, v1, v2}$ from Table B3, Appendix B, GVP	critical value:
Calculations Case B - H_a: $\sigma^2_1 > \sigma^2_2$ or H_a: $\sigma^2_1 < \sigma^2_2$	
Obtain $F_{0.95, v1, v2}$ from Table B3, Appendix B, GVP	critical value:

Decision Rule

If the test statistic F ≥ the critical value we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis $\sigma_1^2 = \sigma_2^2$:	o **Not Rejected**	o **Rejected**
Alternative Hypothesis:	o **Accepted**	o **Not Accepted**

[9] A non-rigorous definition of independence is in section 4.4.
[10] The independence of data sets is defined in section 4.4.

5.4 SAW # 4 Testing Percentage Reduction

H_o: $\mu_2 = (1\text{-}p\%)\mu_1$

This test[11] is used to determine at a level of 95% confidence, whether a pre-specified percentage change occurs in a sample, as the result of applying a process or technology. For example, a claim may state that a technology removes "p%" of contaminant from a process stream (i.e., 95% confident that the technology can remove "p%" of contaminant). If μ_1 is the mean of a sample prior to the application of the technology, we wish to test whether the mean after treatment μ_2 is equal to $(1\text{-}p\%)\mu_1$. The formulae presented below are applicable when the sizes of both data sets are equal or unequal.

Assumptions:

- *Both data sets are normally distributed.*
- *The x_i observations constituting the data set are independent[12].*
- *Data sets are independent of one another[13].*

Data Description	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database

Preliminary Calculations and Tests of Assumptions		
Convert the pre-technology observations x_{1i} for i = 1 … n_1 to $x_{1i}^* = (1\text{-}p\%)\, x_{1i}$.		
Based on SAW#1, samples x_{1i}^* and x_{2i} are normally distributed.	o Yes	o No
If one or both samples are not normally distributed, use SAW #9 to test the equality of median of the transformed pre-technology observations, x_1^* with the median of the post-technology observations, x_2.		
Based on SAW #3, the variances are equal.	o Yes	o No

Common Calculations	
Estimate test statistic, t or Z using SAW #6 if variances are equal or SAW #7 if variances are unequal. Substitute x_{1i}^* for x_{1i} in all calculations.	
Total sample size $n = n_1 + n_2$	n:
Calculations Case A - H_a: $\mu_2 \neq (1\text{-}p\%)\mu_1$	
If n or dof_e[14] $\geq$ 30, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	critical value: 1.960

[11] A more rigorous (but more difficult to implement) test of this hypothesis is provided in Kendall, M. and A. Stuart. 1979. The advanced theory of statistics, Volume 2: Inference and relationship. Chapter 21, pg 152. Charles Griffin and Co. Ltd., London.

[12] A non-rigorous definition of independence is in section 4.4.

[13] The independence of data sets is defined in section 4.4.

[14] For SAW #6, the choice for the use of Z or t is based on n. For SAW #7, the choice for the use of Z or t is based on the effective degrees of freedom, dof_e

If n or dof_e[14] < 30, obtain $t_{0.975,\ n-2\ or\ dofe}$[15] from Table B2, Appendix B, GVP.	critical value:
Calculations Case B $H_a: \mu_2 < (1-p\%)\mu_1$	
If n or dof_e[11] ≥ 30, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645
If n or dof_e[11] < 30, obtain $t_{0.95,\ n-2\ or\ dofe}$[15] from Table B2, Appendix B, GVP	critical value:
Calculations Case C $H_a: \mu_2 > (1-p\%)\mu_1$	
If n or dof_e[11] ≥ 30, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If n or dof_e[11] < 30, obtain $t_{0.95,\ n-2\ or\ dofe}$[12] from Table B2, Appendix B, GVP, and multiply by -1	critical value:

Decision Rule

Inferences Case A:

If the test statistics, |t| or |Z| ≥ critical value we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case B:

If the test statistic, t or Z ≥ critical value we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case C:

If the test statistic, t or Z ≤ critical value we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis: o **Not Rejected** o **Rejected**

Alternative Hypothesis: o **Accepted** o **Not Accepted**

[15] For SAW #6, the degrees of freedom are n-2. For SAW #7, the degrees of freedom are dof_e

5.5 SAW # 5 Testing Mean is Equal to a Specified Value

H_o: $\mu_1 = \mu_o$

This test is used to determine at a level of 95% confidence that the mean is not equal to some pre-specified value, μ_o. *The value μ_o will often be the performance that a technology is claiming to achieve.*

Assumptions:

- *Data set is normally distributed.*
- *The x_i observations constituting the data set are independent*[16].

Data Description and Tests of Assumptions	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database
Based on SAW #1, the data set is normally distributed.	o Yes

Common Calculations	
Estimate of μ	$\bar{x}$:
Hypothesized value μ_o	μ_o:
Sample size n	n:
Estimate of σ^2 $s^2=\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right)$	s^2:
If n < 30, the test statistic t, is given by: $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	t :
If n ≥ 30, the test statistic Z, is given by: $Z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	Z :

Calculations Case A - H_a : $\mu_1 \neq \mu_o$	
If n ≥30, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	critical value:1.960

[16] A non-rigorous definition of independence is in section 4.4.

If n <30, obtain $t_{0.975,\ n-1}$ from Table B2, Appendix B, GVP.	critical value:

Calculations Case B - $H_a: \mu_1 < \mu_o$	
If n ≥30, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If n <30, obtain $t_{0.95,\ n-1}$ from Table B2, Appendix B, GVP, and multiply by -1.	critical value:
Calculations Case C - $H_a: \mu_1 > \mu_o$	
If n ≥30, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645
If n <30, obtain $t_{0.95,\ n-1}$ from Table B2, Appendix B, GVP.	critical value:

Decision Rule

Inferences Case A:

If the test statistics, |t| or |Z| ≥ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case B:

If the test statistics, t or Z ≤ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case C:

If the test statistic, t or Z ≥ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis: o **Not Rejected** o **Rejected**

Alternative Hypothesis: o **Accepted** o **Not Accepted**

5.6 SAW # 6 Testing Equality of Two Means when Sample Variances are Assumed Equal

$$H_o: \mu_1 - \mu_2 = d_o$$

This test is used to determine at a level of 95% confidence, if the difference of two means is equal to a pre-specified value d_o. If this value is 0, we are testing that the two means are equal or, $\mu_1 = \mu_2$. The formulae presented below are applicable when the two data sets are equal or unequal in number.

Assumptions:

- *Both data sets are normally distributed.*
- *Variances estimated from both data sets are equal.*
- *The x_i observations constituting the data set are independent[17].*
- *Data sets are independent of one another[18].*

Data Description and Tests of Assumptions		
Parameter:	Units:	
Data Location	o attached page	
Filename and Location	o electronic database	
Based on SAW#1,the data sets are normally distributed.	o Yes	
Based on SAW #3, the variances are equal.	o Yes	o No use SAW # 7

Common Calculations	
Pre-specified value d_o	d_o :
Estimate of μ_1	$\bar{x}_1$:
Estimate of μ_2	$\bar{x}_2$:
n_1:	n_2:
Total sample size $n = n_1 + n_2$	n:
Estimate of σ_1^2 $s_1^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n_1} x_{1i}\right)^2}{n_1}\right]$	s_1^2 :

[17] A non-rigorous definition of independence is in section 4.4.
[18] The independence of data sets is defined in section 4.4.

Common Calculations	
Estimate of σ_2^2 $s_2^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n_2} x_{2i}\right)^2}{n_2}\right]$	s_2^2:
Estimate of pooled variance σ_p^2 $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	s_p^2:
If $n_1 + n_2 - 2 < 30$, the test statistic t, is given by: $t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t:
If $n_1 + n_2 - 2 \geq 30$, the test statistic Z, is given by: $Z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	Z:
Calculations Case A - H_a: $\mu_1 \neq \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	Critical value:1.960
If $n < 30$, obtain $t_{0.975,\ n-2}$ from Table B2, Appendix B, GVP.	critical value:
Calculations Case B - H_a :$\mu_1 < \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If $n < 30$, obtain $t_{0.95,\ n-2}$ from Table B2, Appendix B, GVP, and multiply by -1.	critical value:
Calculations Case C - H_a :$\mu_1 > \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645
If $n < 30$, obtain $t_{0.95,\ n-2}$ from Table B2, Appendix B, GVP.	critical value:

Decision Rule

Inferences Case A:

If the test statistic, |t| or |Z| ≥ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case B:

If the test statistic, t or Z ≤ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case C:

If the test statistic, t or Z ≥ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis: o **Not Rejected** o **Rejected**

Alternative Hypothesis: o **Accepted** o **Not Accepted**

5.7 SAW # 7 Testing Equality of Two Means when Sample Variances are Assumed Unequal

H_o: $\mu_1 - \mu_2 = d_o$

This test is used to determine at a level of 95% confidence, if the difference of two means is equal to a pre-specified value d_o. If this value is 0, we are testing that the two means are equal or, $\mu_1 = \mu_2$. The formulae presented below are applicable when the two sample sizes are equal or unequal.

Assumptions:

- *Both data sets are normally distributed.*
- *Variances estimated from both data sets are unequal.*
- *The x_i observations constituting the data set are independent[19].*
- *Data sets are independent of one another[20].*

Data Description and Tests of Assumptions	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database
Based on SAW#1,the data sets are normally distributed.	o Yes
Based on SAW #3, the variances are not equal.	o Yes

Common Calculations	
Pre-specified value d_o	d_o:
Estimate of μ_1	$\bar{x}_1$:
Estimate of μ_2	$\bar{x}_2$:
n_1:	n_2:
Total sample size $n = n_1 + n_2$	n:
Estimate of σ_1^2 $s_1^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n_1} x_{1i}\right)^2}{n_1}\right]$	s_1^2:

[19] A non-rigorous definition of independence is in section 4.4.
[20] The independence of data sets is defined in section 4.4.

Common Calculations	
Estimate of σ_2^2 $s_2^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n_2} x_{2i}\right)^2}{n_2}\right]$	s_2^2:
Estimate of effective degrees of freedom (dof_e)[21] $dof_e = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\left[\left(s_1^2/n_1\right)^2/(n_1-1)\right] + \left[\left(s_2^2/n_2\right)^2/(n_2-1)\right]}$	dof_e:
Estimate of pooled standard deviation $s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$s_{\bar{x}_1-\bar{x}_2}$:
If dof_e < 30, the test statistic t, is given by: $t^* = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_{\bar{x}_1-\bar{x}_2}}$	t^* :
If $dof_e \geq 30$, the test statistic Z, is given by: $Z^* = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_{\bar{x}_1-\bar{x}_2}}$	Z^* :
$\alpha = 0.05$	$1-\alpha/2 = 0.975$
Calculations Case A - H_a: $\mu_1 \neq \mu_2 + d_o$	
If $dof_e \geq 30$, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	critical value: 1.960
If dof_e <30, obtain $t_{0.975,\ dofe}$ from Table B2, Appendix B, GVP.	critical value:
Calculations Case B - H_a :$\mu_1 < \mu_2 + d_o$	
If $dof_e \geq 30$, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If dof_e <30, obtain $t_{0.95,\ dofe}$ from Table B2, Appendix B, GVP and multiply by -1.	critical value:
Calculations Case C - H_a :$\mu_1 > \mu_2 + d_o$	
If $dof_e \geq 30$, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645

[21] If the effective degrees of freedom are not a whole number, round the value down to the nearest integer to conserve the stated α value.

Common Calculations	
If dof_e < 30, obtain $t_{0.95,\ dofe}$ from Table B2, Appendix B, GVP.	critical value:

Decision Rule

Inferences Case A:

If the test statistic, $|t^*|$ or $|Z^*| \geq$ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case B:

If the test statistic, t^* or $Z^* \leq$ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case C:

If the test statistic, t^* or $Z^* \geq$ critical value, we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis: o **Not Rejected** o **Rejected**

Alternative Hypothesis: o **Accepted** o **Not Accepted**

5.8 SAW # 8 Testing Median is Equal to a Specified Value
H_o: median = m_o

This test is used to determine at a level of 95% confidence that the median is not equal to some pre-specified value, m_o. *The value, m_o will often be the performance that a technology is claiming to achieve. The test presented is the Wilcoxon Signed Ranks test.*

Assumptions:

- *The x_i observations constituting the data set are independent*[22].
- *The distribution of each d_i is symmetric*[23].

Data Description and Tests of Assumptions	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database
From SAW#1, the data set is not normally distributed.	o True
From a frequency histogram for visual assessment, the data is symmetric	o True

Common Calculations	
Pre-specified value, m_o.	m_o:
Sample size n	n:
Sort the x_i from smallest to largest.	
Calculate the vector d_i $d_i = m_o - x_i$	
Rank the $\|d_i\|$ from smallest to largest to obtain a vector R_i of length n. Identical $\|d_i\|$ are assigned the average of the ranks they would otherwise have received.	
The test statistic T^+ is $T^+ = \sum_{i=1}^{n} R_i$, for positive d_i only	T^+:
Calculations Case A - H_a: median ≠ m_o	
Obtain $w_{0.025}$ from Table B4, Appendix B, GVP.	critical value $w_{0.025}$:
critical value = $w_{0.975} = n(n+1)/2 - w_{0.025}$	critical value $w_{0.975}$:
Calculations Case B - H_a : median < m_o	
Obtain $w_{0.05}$ from Table B4, Appendix B, GVP.	$w_{0.05}$:
critical value = $w_{0.950} = n(n+1)/2 - w_{0.05}$	critical value $w_{0.95}$:
Calculations Case C - H_a : median > m_o	
Obtain $w_{0.05}$ from Table B4, Appendix B, GVP.	critical value $w_{0.05}$:

[22] A non-rigorous definition of independence is in section 4.4.
[23] Symmetric distribution is defined in Appendix D.

Decision Rule

Inference Case A:

If $T^+ < w_{0.025}$ or if $T^+ > w_{0.975}$ we reject the null hypothesis and accept the alternative hypothesis. The median is not equal to m_o.

Inference Case B:

If $T^+ \geq w_{0.95}$ we reject the null hypothesis and accept the alternative hypothesis. The median is $< m_o$.

Inference Case C:

If $T^+ \leq w_{0.05}$ we reject the null hypothesis and accept the alternative hypothesis. The median is $> m_o$.

Null Hypothesis: o **Not Rejected** o **Rejected**

Alternative Hypothesis: o **Accepted** o **Not Accepted**

5.9 SAW # 9Testing Equality of Two Medians

H_o: $median_1$ - $median_2$ = d_o

This test is used to determine at a level of 95% confidence, if the difference between two medians is equal to a pre-specified value d_o. If this value is 0, we are testing that the two medians are equal, or $median_1$ = $median_2$. The formulae presented below are applicable when the two sample sizes are equal or unequal and do not assume any distribution form. The Mann-Whitney test uses the ranks of the data sets to test hypotheses.

Assumptions:

- *Data sets are independent on one another[24].*
- *The x_i observations constituting the data set are independent[25].*

Data Description and Tests of Assumptions	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database
From SAW#1, the data sets are not normally distributed.	o Yes

Common Calculations	
Pre-specified value, d_o.	d_o:
Sample size of group 1.	n_1:
Sample size of group 2.	n_2:
Total sample size $n = n_1 + n_2$	n:

Test Procedure	
Rank the combined observations from groups 1 and 2 so that the ranks $R(x_i)$, range from 1 to n. In the case of identical or tied ranks, assign the average of the ranks that otherwise would have been assigned.	
Test statistic T $T = \sum_{i=1}^{n_1} R(X_i)$, where X_i are the n_1 observations from group 1.	T:
Calculations Case A - H_a: $median_1 \neq median_2 + d_o$	
Obtain $w_{0.025}$ from Table B7, Appendix B, GVP.	critical value $w_{0.025}$:
$w_{0.975} = n_1(n + 1) - w_{0.025}$	critical value $w_{0.975}$:
Calculations Case B - H_a :$median_1 < median_2 + d_o$	
Obtain $w_{0.05}$ from Table B4, Appendix B, GVP.	critical value $w_{0.05}$:

[24] The independence of two data sets is defined in section 4.4.
[25] A non-rigorous definition of independence is in section 4.4.

Calculations Case C - H_a :median$_1$ > median$_2$ + d_o	
Obtain value of $w_{0.05}$ from Table B4, Appendix B, GVP.	$w_{0.05}$:
$w_{0.95} = n_1(n + 1) - w_{0.05}$	critical value $w_{0.95}$:

Decision Rule

Inferences Case A:

If $T < w_{0.025}$ or if $T > w_{0.975}$ we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case B:

If $T < w_{0.05}$ we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case C:

If $T > w_{0.95}$ we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis: o **Not Rejected** o **Rejected**

Alternative Hypothesis: o **Accepted** o **Not Accepted**

5.10 SAW # 10 Testing For Mean Difference in Paired Observations

H_o: $\mu_1 - \mu_2 = d_o$

This test is used to determine at a level of 95% confidence, if the difference of means of two paired variables is equal to a pre-specified value, d_o. If this value is 0, we are testing that the two means are equal or, $\mu_1 = \mu_2$.

Assumptions:

- *The vector of differences[26] is normally distributed.*
- *Data sets are of the same size.*
- *The x_i observations constituting the data set are independent[27].*

Data Description	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database

Pre-Test Calculations and Tests of Assumptions		
Calculate the difference $d_i = x_{1i} - x_{2i}$ for each pair of observations. This will produce a vector of n observations. Perform subsequent calculations on the difference d_i.		
Based on SAW #1, the vector of differences is normally distributed.	o Yes Continue.	o No See below.
If data is non-normal, the median of the vector of differences may be used to test H_o: $\mu_1 - \mu_2 = d_o$, by testing H_o: median of $d_i = d_o$, using SAW # 8.		

Common Calculations	
Pre-specified value, d_o.	d_o:
Sample size n is the number of differences.	n:
Estimate of μ_d $\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$	$\bar{d}$:

[26] The difference is defined in "Pre-Test Calculations and Tests of Assumptions".
[27] A non-rigorous definition of independence is in section 4.4.

Common Calculations	
Estimate of σ^2 $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n}\right]$	s^2 :
If n < 30, the test statistic t, is given by: $t = \frac{\bar{d} - d_0}{s/\sqrt{n}}$	t :
If n ≥ 30, the test statistic Z, is given by: $Z = \frac{\bar{d} - d_0}{s/\sqrt{n}}$	Z :
Calculations Case A - H_a: $\mu_1 \neq \mu_2 + d_o$	
If n ≥ 30, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	critical value:1.960
If n <30, obtain $t_{0.975,\ n-1}$ from Table B2, Appendix B, GVP.	critical value:
Calculations Case B - H_a :$\mu_1 < \mu_2 + d_o$	
If n ≥ 30, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If n <30, obtain $t_{0.95,\ n-1}$ from Table B2, Appendix B, GVP, and multiply by –1.	critical value:
Calculations Case C - H_a :$\mu_1 > \mu_2 + d_o$	
If n ≥ 30, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645
If n <30, obtain $t_{0.95,\ n-1}$ from Table B2, Appendix B, GVP.	critical value:

Decision Rule

Inferences Case A:

If the test statistics, |t| or |Z| ≥ critical value we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case B:

If the test statistics, t or Z ≤ critical value we reject the null hypothesis and accept the alternative hypothesis.

Inferences Case C:

If the test statistics, t or Z ≥ critical value we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis: o **Not Rejected** o **Rejected**

Alternative Hypothesis: o **Accepted** o **Not Accepted**

Environmental Technology Verification

General Verification Protocol

Appendix B

Statistical Tables

February 2007
(Rev. May 2013)

Environment Canada Environnement Canada

LIST OF TABLES

Table B1. Area under a Normal Distribution 2
Table B2. Percentage Points of the t Distribution 5
Table B3. Values of $F_{0.95}$* (SAW #3) 6
Table B4. Quartiles of the Wilcoxon Signed-Rank Test* (SAW #8) 8
Table B5. Quartiles of the Mann-Whitney Test Statistic* (SAW# 9) 10

Table B1. Area under a Normal Distribution

$$Z = \frac{\bar{x} - \mu}{\sigma / n^{1/2}}$$

Z	P	Z	P	Z	P	Z	P	Z	P	Z	P
-3.50	0.0002	-3.05	0.0011	-2.60	0.0047	-2.15	0.0158	-1.70	0.0446	-1.25	0.1056
-3.49	0.0002	-3.04	0.0012	-2.59	0.0048	-2.14	0.0162	-1.69	0.0455	-1.24	0.1075
-3.48	0.0003	-3.03	0.0012	-2.58	0.0049	-2.13	0.0166	-1.68	0.0465	-1.23	0.1093
-3.47	0.0003	-3.02	0.0013	-2.57	0.0051	-2.12	0.0170	-1.67	0.0475	-1.22	0.1112
-3.46	0.0003	-3.01	0.0013	-2.56	0.0052	-2.11	0.0174	-1.66	0.0485	-1.21	0.1131
-3.45	0.0003	-3.00	0.0013	-2.55	0.0054	-2.10	0.0179	-1.65	0.0495	-1.20	0.1151
-3.44	0.0003	-2.99	0.0014	-2.54	0.0055	-2.09	0.0183	-1.64	0.0505	-1.19	0.1170
-3.43	0.0003	-2.98	0.0014	-2.53	0.0057	-2.08	0.0188	-1.63	0.0516	-1.18	0.1190
-3.42	0.0003	-2.97	0.0015	-2.52	0.0059	-2.07	0.0192	-1.62	0.0526	-1.17	0.1210
-3.41	0.0003	-2.96	0.0015	-2.51	0.0060	-2.06	0.0197	-1.61	0.0537	-1.16	0.1230
-3.40	0.0003	-2.95	0.0016	-2.50	0.0062	-2.05	0.0202	-1.60	0.0548	-1.15	0.1251
-3.39	0.0003	-2.94	0.0016	-2.49	0.0064	-2.04	0.0207	-1.59	0.0559	-1.14	0.1271
-3.38	0.0004	-2.93	0.0017	-2.48	0.0066	-2.03	0.0212	-1.58	0.0571	-1.13	0.1292
-3.37	0.0004	-2.92	0.0018	-2.47	0.0068	-2.02	0.0217	-1.57	0.0582	-1.12	0.1314
-3.36	0.0004	-2.91	0.0018	-2.46	0.0069	-2.01	0.0222	-1.56	0.0594	-1.11	0.1335
-3.35	0.0004	-2.90	0.0019	-2.45	0.0071	-2.00	0.0228	-1.55	0.0606	-1.10	0.1357
-3.34	0.0004	-2.89	0.0019	-2.44	0.0073	-1.99	0.0233	-1.54	0.0618	-1.09	0.1379
-3.33	0.0004	-2.88	0.0020	-2.43	0.0075	-1.98	0.0239	-1.53	0.0630	-1.08	0.1401
-3.32	0.0005	-2.87	0.0021	-2.42	0.0078	-1.97	0.0244	-1.52	0.0643	-1.07	0.1423
-3.31	0.0005	-2.86	0.0021	-2.41	0.0080	-1.96	0.0250	-1.51	0.0655	-1.06	0.1446
-3.30	0.0005	-2.85	0.0022	-2.40	0.0082	-1.95	0.0256	-1.50	0.0668	-1.05	0.1469
-3.29	0.0005	-2.84	0.0023	-2.39	0.0084	-1.94	0.0262	-1.49	0.0681	-1.04	0.1492
-3.28	0.0005	-2.83	0.0023	-2.38	0.0087	-1.93	0.0268	-1.48	0.0694	-1.03	0.1515
-3.27	0.0005	-2.82	0.0024	-2.37	0.0089	-1.92	0.0274	-1.47	0.0708	-1.02	0.1539
-3.26	0.0006	-2.81	0.0025	-2.36	0.0091	-1.91	0.0281	-1.46	0.0721	-1.01	0.1562
-3.25	0.0006	-2.80	0.0026	-2.35	0.0094	-1.90	0.0287	-1.45	0.0735	-1.00	0.1587
-3.24	0.0006	-2.79	0.0026	-2.34	0.0096	-1.89	0.0294	-1.44	0.0749	-0.99	0.1611
-3.23	0.0006	-2.78	0.0027	-2.33	0.0099	-1.88	0.0301	-1.43	0.0764	-0.98	0.1635
-3.22	0.0006	-2.77	0.0028	-2.32	0.0102	-1.87	0.0307	-1.42	0.0778	-0.97	0.1660
-3.21	0.0007	-2.76	0.0029	-2.31	0.0104	-1.86	0.0314	-1.41	0.0793	-0.96	0.1685
-3.20	0.0007	-2.75	0.0030	-2.30	0.0107	-1.85	0.0322	-1.40	0.0808	-0.95	0.1711
-3.19	0.0007	-2.74	0.0031	-2.29	0.0110	-1.84	0.0329	-1.39	0.0823	-0.94	0.1736
-3.18	0.0007	-2.73	0.0032	-2.28	0.0113	-1.83	0.0336	-1.38	0.0838	-0.93	0.1762
-3.17	0.0008	-2.72	0.0033	-2.27	0.0116	-1.82	0.0344	-1.37	0.0853	-0.92	0.1788
-3.16	0.0008	-2.71	0.0034	-2.26	0.0119	-1.81	0.0351	-1.36	0.0869	-0.91	0.1814
-3.15	0.0008	-2.70	0.0035	-2.25	0.0122	-1.80	0.0359	-1.35	0.0885	-0.90	0.1841
-3.14	0.0008	-2.69	0.0036	-2.24	0.0125	-1.79	0.0367	-1.34	0.0901	-0.89	0.1867
-3.13	0.0009	-2.68	0.0037	-2.23	0.0129	-1.78	0.0375	-1.33	0.0918	-0.88	0.1894
-3.12	0.0009	-2.67	0.0038	-2.22	0.0132	-1.77	0.0384	-1.32	0.0934	-0.87	0.1922
-3.11	0.0009	-2.66	0.0039	-2.21	0.0136	-1.76	0.0392	-1.31	0.0951	-0.86	0.1949
-3.10	0.0010	-2.65	0.0040	-2.20	0.0139	-1.75	0.0401	-1.30	0.0968	-0.85	0.1977
-3.09	0.0010	-2.64	0.0041	-2.19	0.0143	-1.74	0.0409	-1.29	0.0985	-0.84	0.2005
-3.08	0.0010	-2.63	0.0043	-2.18	0.0146	-1.73	0.0418	-1.28	0.1003	-0.83	0.2033
-3.07	0.0011	-2.62	0.0044	-2.17	0.0150	-1.72	0.0427	-1.27	0.1020	-0.82	0.2061
-3.06	0.0011	-2.61	0.0045	-2.16	0.0154	-1.71	0.0436	-1.26	0.1038	-0.81	0.2090
-0.80	0.2119	-0.35	0.3632	0.10	0.5398	0.55	0.7088	1.00	0.8413	1.45	0.9265
-0.79	0.2148	-0.34	0.3669	0.11	0.5438	0.56	0.7123	1.01	0.8438	1.46	0..9279
-0.78	0.2177	-0.33	0.3707	0.12	0.5478	0.57	0.7157	1.02	0.8461	1.47	0.9292
-0.77	0.2206	-0.32	0.3745	0.13	0.5517	0.58	0.7190	1.03	0.8485	1.48	0.9306

Table B1. Area under a Normal Distribution

$$Z = \frac{\bar{x} - \mu}{\sigma / n^{1/2}}$$

Z	P	Z	P	Z	P	Z	P	Z	P	Z	P
-0.76	0.2236	-0.31	0.3783	0.14	0.5557	0.59	0.7224	1.04	0.8508	1.49	0.9319
-0.75	0.2266	-0.30	0.3821	0.15	0.5596	0.60	0.7257	1.05	0.8531	1.50	0.9332
-0.74	0.2296	-0.29	0.3859	0.16	0.5636	0.61	0.7291	1.06	0.8554	1.51	0.9345
-0.73	0.2327	-0.28	0.3897	0.17	0.5675	0.62	0.7324	1.07	0.8577	1.52	0.9357
-0.72	0.2358	-0.27	0.3936	0.18	0.5714	0.63	0.7357	1.08	0.8599	1.53	0.9370
-0.71	0.2389	-0.26	0.3974	0.19	0.5753	0.64	0.7389	1.09	0.8621	1.54	0.9382
-0.70	0.2420	-0.25	0.4013	0.20	0.5793	0.65	0.7422	1.10	0.8643	1.55	0.9394
-0.69	0.2451	-0.24	0.4052	0.21	0.5832	0.66	0.7454	1.11	0.8665	1.56	0.9406
-0.68	0.2483	-0.23	0.4090	0.22	0.5871	0.67	0.7486	1.12	0.8686	1.57	0.9418
-0.67	0.2514	-0.22	0.4129	0.23	0.5910	0.68	0.7517	1.13	0.8708	1.58	0.9429
-0.66	0.2546	-0.21	0.4168	0.24	0.5948	0.69	0.7549	1.14	0.8729	1.59	0.9441
-0.65	0.2578	-0.20	0.4207	0.25	0.5987	0.70	0.7580	1.15	0.8749	1.60	0.9452
-0.64	0.2611	-0.19	0.4247	0.26	0.6026	0.71	0.7611	1.16	0.8770	1.61	0.9463
-0.63	0.2643	-0.18	0.4286	0.27	0.6064	0.72	0.7642	1.17	0.8790	1.62	0.9474
-0.62	0.2676	-0.17	0.4325	0.28	0.6103	0.73	0.7673	1.18	0.8810	1.63	0.9484
-0.61	0.2709	-0.16	0.4364	0.29	0.6141	0.74	0.7704	1.19	0.8830	1.64	0.9495
-0.60	0.2743	-0.15	0.4404	0.30	0.6179	0.75	0.7734	1.20	0.8849	1.65	0.9505
-0.59	0.2776	-0.14	0.4443	0.31	0.6217	0.76	0.7764	1.21	0.8869	1.66	0.9515
-0.58	0.2810	-0.13	0.4483	0.32	0.6255	0.77	0.7794	1.22	0.8888	1.67	0.9525
-0.57	0.2843	-0.12	0.4522	0.33	0.6293	0.78	0.7823	1.23	0.8907	1.68	0.9535
-0.56	0.2877	-0.11	0.4562	0.34	0.6331	0.79	0.7852	1.24	0.8925	1.69	0.9545
-0.55	0.2912	-0.10	0.4602	0.35	0.6368	0.80	0.7881	1.25	0.8944	1.70	0.9554
-0.54	0.2946	-0.09	0.4641	0.36	0.6406	0.81	0.7910	1.26	0.8962	1.71	0.9564
-0.53	0.2981	-0.08	0.4681	0.37	0.6443	0.82	0.7939	1.27	0.8980	1.72	0.9573
-0.52	0.3015	-0.07	0.4721	0.38	0.6480	0.83	0.7967	1.28	0.8997	1.73	0.9582
-0.51	0.3050	-0.06	0.4761	0.39	0.6517	0.84	0.7995	1.29	0.9015	1.74	0.9591
-0.50	0.3085	-0.05	0.4801	0.40	0.6554	0.85	0.8023	1.30	0.9032	1.75	0.9599
-0.49	0.3121	-0.04	0.4840	0.41	0.6591	0.86	0.8051	1.31	0.9049	1.76	0.9608
-0.48	0.3156	-0.03	0.4880	0.42	0.6628	0.87	0.8078	1.32	0.9066	1.77	0.9616
-0.47	0.3192	-0.02	0.4920	0.43	0.6664	0.88	0.8106	1.33	0.9082	1.78	0.9625
-0.46	0.3228	-0.01	0.4960	0.44	0.6700	0.89	0.8133	1.34	0.9099	1.79	0.9633
-0.45	0.3264	**0.00**	**0.5000**	0.45	0.6736	0.90	0.8159	1.35	0.9115	1.80	0.9641
-0.44	0.3300	0.01	0.5040	0.46	0.6772	0.91	0.8186	1.36	0.9131	1.81	0.9649
-0.43	0.3336	0.02	0.5080	0.47	0.6808	0.92	0.8212	1.37	0.9147	1.82	0.9656
-0.42	0.3372	0.03	0.5120	0.48	0.6844	0.93	0.8238	1.38	0.9162	1.83	0.9664
-0.41	0.3409	0.04	0.5160	0.49	0.6879	0.94	0.8264	1.39	0.9177	1.84	0.9671
-0.40	0.3446	0.05	0.5199	0.50	0.6915	0.95	0.8289	1.40	0.9192	1.85	0.9678
-0.39	0.3483	0.06	0.5239	0.51	0.6950	0.96	0.8315	1.41	0.9207	1.86	0.9686
-0.38	0.3520	0.07	0.5279	0.52	0.6985	0.97	0.8340	1.42	0.9222	1.87	0.9693
-0.37	0.3557	0.08	0.5319	0.53	0.7019	0.98	0.8365	1.43	0.9236	1.88	0.9699
-0.36	0.3594	0.09	0.5359	0.54	0.7054	0.99	0.8389	1.44	0.9251	1.89	0.9706
1.90	0.9713	2.20	0.9861	2.50	0.9938	2.80	0.9974	3.10	0.9990	3.40	0.9997
1.91	0.9719	2.21	0.9864	2.51	0.9940	2.81	0.9975	3.11	0.9991	3.41	0.9997
1.92	0.9726	2.22	0.9868	2.52	0.9941	2.82	0.9976	3.12	0.9991	3.42	0.9997
1.93	0.9732	2.23	0.9871	2.53	0.9943	2.83	0.9977	3.13	0.9991	3.43	0.9997
1.94	0.9738	2.24	0.9875	2.54	0.9945	2.84	0.9977	3.14	0.9992	3.44	0.9997
1.95	0.9744	2.25	0.9878	2.55	0.9946	2.85	0.9978	3.15	0.9992	3.45	0.9997
1.96	0.9750	2.26	0.9881	2.56	0.9948	2.86	0.9979	3.16	0.9992	3.46	0.9997
1.97	0.9756	2.27	0.9884	2.57	0.9949	2.87	0.9979	3.17	0.9992	3.47	0.9997

Table B1. Area under a Normal Distribution

$$Z = \frac{\bar{x} - \mu}{\sigma / n^{1/2}}$$

Z	P	Z	P	Z	P	Z	P	Z	P	Z	P
1.98	0.9761	2.28	0.9887	2.58	0.9951	2.88	0.9980	3.18	0.9993	3.48	0.9997
1.99	0.9767	2.29	0.9890	2.59	0.9952	2.89	0.9981	3.19	0.9993	3.49	0.9998
2.00	0.9772	2.30	0.9893	2.60	0.9953	2.90	0.9981	3.20	0.9993	3.50	0.9998
2.01	0.9778	2.31	0.9896	2.61	0.9955	2.91	0.9982	3.21	0.9993		
2.02	0.9783	2.32	0.9898	2.62	0.9956	2.92	0.9982	3.22	0.9994		
2.03	0.9788	2.33	0.9901	2.63	0.9957	2.93	0.9983	3.23	0.9994		
2.04	0.9793	2.34	0.9904	2.64	0.9959	2.94	0.9984	3.24	0.9994		
2.05	0.9798	2.35	0.9906	2.65	0.9960	2.95	0.9984	3.25	0.9994		
2.06	0.9803	2.36	0.9909	2.66	0.9961	2.96	0.9985	3.26	0.9994		
2.07	0.9808	2.37	0.9911	2.67	0.9962	2.97	0.9985	3.27	0.9995		
2.08	0.9812	2.38	0.9913	2.68	0.9963	2.98	0.9986	3.28	0.9995		
2.09	0.9817	2.39	0.9916	2.69	0.9964	2.99	0.9986	3.29	0.9995		
2.10	0.9821	2.40	0.9918	2.70	0.9965	3.00	0.9987	3.30	0.9995		
2.11	0.9826	2.41	0.9920	2.71	0.9966	3.01	0.9987	3.31	0.9995		
2.12	0.9830	2.42	0.9922	2.72	0.9967	3.02	0.9987	3.32	0.9995		
2.13	0.9834	2.43	0.9925	2.73	0.9968	3.03	0.9988	3.33	0.9996		
2.14	0.9838	2.44	0.9927	2.74	0.9969	3.04	0.9988	3.34	0.9996		
2.15	0.9842	2.45	0.9929	2.75	0.9970	3.05	0.9989	3.35	0.9996		
2.16	0.9846	2.46	0.9931	2.76	0.9971	3.06	0.9989	3.36	0.9996		
2.17	0.9850	2.47	0.9932	2.77	0.9972	3.07	0.9989	3.37	0.9996		
2.18	0.9854	2.48	0.9934	2.78	0.9973	3.08	0.9990	3.38	0.9996		
2.19	0.9857	2.49	0.9936	2.79	0.9974	3.09	0.9990	3.39	0.9997		

Table B2. Percentage Points of the t Distribution

df*	$t_{0.6}$	$t_{0.7}$	$t_{0.8}$	$t_{0.9}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.3249	0.7265	1.3764	3.0777	6.3138	12.706	31.821	63.657
2	0.2887	0.6172	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.2767	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.2707	0.5686	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.2672	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.2648	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.2632	0.5491	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.2619	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.2610	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.2602	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.2596	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.2590	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.2586	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.2582	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.2579	0.5357	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467
16	0.2576	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.2573	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.2571	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.2569	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.2567	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.2566	0.5325	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.2564	0.5321	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.2563	0.5317	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.2562	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.2561	0.5312	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.2560	0.5309	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.2559	0.5306	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.2558	0.5304	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.2557	0.5302	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.2556	0.5300	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.2555	0.5298	0.8534	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.2555	0.5297	0.8530	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.2554	0.5295	0.8526	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.2553	0.5294	0.8523	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.2553	0.5292	0.8520	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.2552	0.5291	0.8517	1.3055	1.6883	2.0281	2.4345	2.7195

* df – degrees of freedom

Table B3. Values of $F_{0.95}$* (SAW #3)

ν_2 - Degrees of freedom for denominator	ν_1 = Degrees of freedom for numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.38	2.38	2.30	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	3.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.93
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

* This table is reproduced from Miller and Freund, “Probability and Statistics for Engineers”.

Table B3. Values of $F_{0.975}$* (SAW #3)

ν_2 - Degrees of freedom for denominator	ν_1 = Degrees of freedom for numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	648	800	864	890	923	937	948	957	963	969	977	985	993	997	1001	1006	1010	1014	1018
2	38.5	39.0	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5	39.5	39.5
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.3	14.2	14.1	14.1	14.0	14.0	14.0	13.9
4	12.2	10.7	10.0	9.60	9.36	9.20	9.07	8.98	8.90	8.75	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	5.70	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.67	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.5	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

Table B4. Quartiles of the Wilcoxon Signed-Rank Test* (SAW #8)

n	α									n(n+1)/2
	0.005	0.010	0.025	0.050	0.100	0.200	0.300	0.400	0.500	
4	0	0	0	0	1	3	3	4	5	10
5	0	0	0	1	3	4	5	6	7.50	15
6	0	0	1	3	4	6	8	9	10.50	21
7	0	1	3	4	6	9	11	12	14	28
8	1	2	4	6	9	12	14	16	18	36
9	2	4	6	9	11	15	18	20	22.50	45
10	4	6	9	11	15	19	22	25	27.50	55
11	6	8	11	14	18	23	27	30	33	66
12	8	10	14	18	22	28	32	36	39	78
13	10	13	18	22	27	33	38	42	45.50	91
14	13	16	22	26	32	39	44	48	52.50	105
15	16	20	26	31	37	45	51	55	60	120
16	20	24	30	36	43	51	58	63	68	136
17	24	28	35	42	49	58	65	71	76.50	153
18	28	33	41	48	56	66	73	80	85.50	171
19	33	38	47	54	63	74	82	89	95	190
20	38	44	53	61	70	83	91	98	105	210
21	44	50	59	68	78	91	100	108	115.50	231
22	49	56	67	76	87	100	110	119	126.50	253
23	55	63	74	84	95	110	120	130	138	276
24	62	70	82	92	105	120	131	141	150	300
25	69	77	90	101	114	131	143	153	162.50	325
26	76	85	99	111	125	142	155	165	175.50	351
27	84	94	108	120	135	154	167	178	189	378
28	92	102	117	131	146	166	180	192	203	406
29	101	111	127	141	158	178	193	206	217.50	435
30	110	121	138	152	170	191	207	220	232.50	465
31	119	131	148	164	182	205	221	235	248	496
32	129	141	160	176	195	219	236	250	264	528
33	139	152	171	188	208	233	251	266	280.50	561
34	149	163	183	201	222	248	266	282	297.50	595
35	160	175	196	214	236	263	283	299	315	630
36	172	187	209	228	251	279	299	317	333	666
37	184	199	222	242	266	295	316	335	351.50	703
38	196	212	236	257	282	312	334	353	370.50	741
39	208	225	250	272	298	329	352	372	390	780
40	221	239	265	287	314	347	371	391	410	820
41	235	253	280	303	331	365	390	411	430.50	861
42	248	267	295	320	349	384	409	431	451.50	903
43	263	282	311	337	366	403	429	452	473	946
44	277	297	328	354	385	422	450	473	495	990
45	292	313	344	372	403	442	471	495	517.50	1035
46	308	329	362	390	423	463	492	517	540.50	1081
47	324	346	379	408	442	484	514	540	564	1128
48	340	363	397	428	463	505	536	563	588	1176
49	357	381	416	447	483	527	559	587	612.50	1225
50	374	398	435	467	504	550	583	611	637.50	1275

* Table entries are w_α of the Wilcoxon Signed Rank test statistic T^+, where $P(T^+ < w_\alpha) \leq \alpha$.

If n > 50 then approximate w_α by:

$w_\alpha = n(n+1)/4 + Z_\alpha \sqrt{n(n+1)(2n+1)/24}$, where Z_α is the standard normal quartile.

Quartiles of w_α for $\alpha > 0.5$ may be calculated by:

$w_\alpha = n(n + 1) / 2 – w_{1-\alpha}$

This page intentionally left Blank

Table B5. Quartiles of the Mann-Whitney Test Statistic* (SAW# 9)

α	n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.001		3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.005		3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4
0.010	**2**	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5
0.025		3	3	3	3	3	3	4	4	4	5	5	5	5	5	5	6	6	6	6
0.050		3	3	3	4	4	4	5	5	5	5	6	6	7	7	7	7	8	8	8
0.100		3	4	4	5	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11
0.001		6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7
0.005		6	6	6	6	6	6	6	7	7	7	8	8	8	9	9	9	9	10	10
0.010	**3**	6	6	6	6	6	7	7	8	8	8	9	9	9	10	10	11	11	11	12
0.025		6	6	6	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15
0.050		6	7	7	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17	18
0.100		7	8	8	9	10	11	12	12	13	14	15	16	17	17	18	19	20	21	22
0.001		10	10	10	10	10	10	10	10	11	11	11	12	12	12	13	13	14	14	14
0.005		10	10	10	10	11	11	12	12	13	13	14	14	15	16	16	17	17	18	19
0.010	**4**	10	10	10	11	12	12	13	14	14	15	16	16	17	18	18	19	20	20	21
0.025		10	10	11	12	13	14	15	15	16	17	18	19	20	21	22	22	23	24	25
0.050		10	11	12	13	14	15	16	17	18	19	20	21	22	23	25	26	27	28	29
0.100		11	12	14	15	16	17	18	20	21	22	23	24	26	27	28	29	31	32	33
0.001		15	15	15	15	15	15	16	17	17	18	18	19	19	20	21	21	22	23	23
0.005		15	15	15	16	17	17	18	19	20	21	22	23	23	24	25	26	27	28	29
0.010	**5**	15	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
0.025		15	16	17	18	19	21	22	23	24	25	27	28	29	30	31	33	34	35	36
0.050		16	17	18	20	21	22	24	25	27	28	29	31	32	34	35	36	38	39	41
0.100		17	18	20	21	23	24	26	28	29	31	33	34	36	38	39	41	43	44	46

α	n_2 / n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.001		21	21	21	21	21	21	23	24	25	26	26	27	28	29	30	31	32	33	34
0.005		21	21	22	23	24	25	26	27	28	29	31	32	33	34	35	37	38	39	40
0.010	**6**	21	21	23	24	25	26	28	29	30	31	33	34	35	37	38	40	41	42	44
0.025		21	23	24	25	27	28	30	32	33	35	36	38	39	41	43	44	46	47	49
0.050		22	24	25	27	29	30	32	34	36	38	39	41	43	45	47	48	50	52	54
0.100		23	25	27	29	31	33	35	37	39	41	43	45	47	49	51	53	56	58	60
0.001		28	28	28	28	29	30	31	32	34	35	36	37	38	39	40	42	43	44	45
0.005		28	28	29	30	32	33	35	36	38	39	41	42	44	45	47	48	50	51	53
0.010	**7**	28	29	30	32	33	35	36	38	40	41	43	45	46	48	50	52	53	55	57
0.025		28	30	32	34	35	37	39	41	43	45	47	49	51	53	55	57	59	61	63
0.050		29	31	33	35	37	40	42	44	46	48	50	53	55	57	59	62	64	66	68
0.100		30	33	35	37	40	42	45	47	50	52	55	57	60	62	65	67	70	72	75
0.001		36	36	36	37	38	39	41	42	43	45	46	48	49	51	52	54	55	57	58
0.005		36	36	38	39	41	43	744	46	48	50	52	54	55	57	59	61	63	65	67
0.010	**8**	36	37	39	41	43	44	46	48	50	52	54	56	59	61	63	65	67	69	71
0.025		37	39	41	43	45	47	50	52	54	56	59	61	63	66	68	71	73	75	78
0.050		38	40	42	45	47	50	52	55	57	60	63	65	68	70	73	76	78	81	84
0.100		39	42	44	47	50	53	56	59	61	64	67	70	73	76	79	82	85	88	91
0.001		45	45	45	47	48	49	51	53	54	56	58	60	61	63	65	67	69	71	72
0.005		45	46	47	49	51	53	55	57	59	62	64	66	68	70	73	75	77	79	82
0.010	**9**	45	47	49	51	53	55	57	60	62	64	67	69	72	74	77	79	82	84	86
0.025		46	48	50	53	56	58	61	63	66	69	72	74	77	80	83	85	88	91	94
0.050		47	50	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100
0.100		48	51	55	58	61	64	68	71	74	77	81	84	87	91	94	98	101	104	108

α	n_2 / n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.001		55	55	56	57	59	61	62	64	66	68	70	73	75	77	79	81	83	85	88
0.005		55	56	58	60	62	65	67	69	72	74	77	80	82	85	87	90	93	95	98
0.010	**10**	55	57	59	62	64	67	69	72	75	78	80	83	86	89	92	94	97	100	103
0.025		56	59	61	64	67	70	73	76	79	82	85	89	92	95	98	101	104	108	111
0.050		57	60	63	67	70	73	76	80	83	87	90	93	97	100	104	107	111	114	118
0.100		59	62	66	69	73	77	80	84	88	92	95	99	103	107	110	114	118	122	126
0.001		66	66	67	69	71	73	75	77	79	82	84	87	89	91	94	96	99	101	104
0.005		66	67	69	72	74	77	80	83	85	88	91	94	97	100	103	106	109	112	115
0.010	**11**	66	68	71	74	76	79	82	85	89	92	95	98	101	104	108	111	114	117	120
0.025		67	70	73	76	80	83	86	90	93	97	100	104	107	111	114	118	122	125	129
0.050		68	72	75	79	83	86	90	94	98	101	105	109	113	117	121	124	128	132	136
0.100		70	74	78	82	86	90	94	98	103	107	111	115	119	124	128	132	136	140	145
0.001		78	78	79	81	83	86	88	91	93	96	98	102	104	106	110	113	116	118	121
0.005		78	80	82	85	88	91	94	97	100	103	106	110	113	116	120	123	126	130	133
0.010	**12**	78	81	84	87	90	93	96	100	103	107	110	114	117	121	125	128	132	135	139
0.025		80	83	86	90	93	97	101	105	108	112	116	120	124	128	132	136	140	144	148
0.050		81	84	88	92	96	100	105	109	111	117	121	126	130	134	139	143	147	151	156
0.100		83	87	91	96	100	105	109	114	118	123	128	132	137	142	146	151	156	160	165
0.001		91	91	93	95	97	100	103	106	109	112	115	118	121	124	127	130	134	137	140
0.005		91	93	95	99	102	105	109	112	116	119	123	126	130	134	137	141	145	149	152
0.010	**13**	92	94	97	101	104	108	112	115	119	123	127	131	135	139	143	147	151	155	159
0.025		93	96	100	104	108	112	116	120	125	129	133	137	142	146	151	155	159	164	168
0.050		94	98	102	107	111	116	120	125	129	134	139	143	148	153	157	162	167	172	176
0.100		96	101	105	110	115	120	125	130	135	140	145	150	155	160	166	171	176	181	186
0.001		105	105	107	109	112	115	118	121	125	128	131	135	138	142	145	149	152	156	160
0.005		105	107	110	113	117	121	124	128	132	136	140	144	148	152	156	160	164	169	173

α	n_2 / n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.010	**14**	106	108	112	116	119	123	128	132	136	140	144	149	153	157	162	166	171	175	179
0.025		107	111	115	119	123	128	132	137	142	146	151	156	161	165	170	175	180	184	189
0.050		109	113	117	122	127	132	137	142	147	152	157	162	167	172	177	183	188	193	198
0.100		110	116	121	126	131	137	142	147	153	158	164	169	175	180	186	191	197	203	208
0.001		120	120	122	125	128	133	135	138	142	145	149	153	157	161	164	168	172	176	180
0.005		120	123	126	129	133	137	141	145	150	154	158	163	167	172	176	181	185	190	194
0.010	**15**	121	124	128	132	136	140	145	149	154	158	163	168	172	177	182	187	191	196	201
0.025		122	126	131	135	140	145	150	155	160	165	170	175	180	185	191	196	201	206	211
0.050		124	128	133	139	144	149	154	160	165	171	176	182	187	193	198	204	209	215	221
0.100		126	131	137	143	148	154	160	166	172	178	184	189	195	201	207	213	219	225	231
0.001		136	136	139	142	145	148	152	156	160	164	168	172	176	180	185	189	193	197	202
0.005		136	139	142	146	150	155	159	164	168	173	178	182	187	192	197	202	207	211	216
0.010	**16**	137	140	144	149	153	158	163	168	173	178	183	188	193	198	203	208	213	219	224
0.025		138	143	148	152	158	163	168	174	179	184	190	196	201	207	212	218	223	229	235
0.050		140	145	151	156	162	167	173	179	185	191	197	202	208	214	220	226	232	238	244
0.100		142	148	154	160	166	173	179	185	191	198	204	211	217	223	230	236	243	249	256
0.001		153	154	156	159	163	167	171	175	179	183	188	192	197	201	206	211	215	220	224
0.005		153	156	160	164	169	173	178	183	188	193	198	203	208	214	219	224	229	235	240
0.010	**17**	154	158	162	167	172	177	182	187	192	198	203	209	214	220	225	231	236	242	247
0.025		156	160	165	171	176	182	188	193	199	205	211	217	223	229	235	241	247	253	259
0.050		157	163	169	174	180	187	193	199	205	211	218	224	231	237	243	250	256	263	269
0.100		160	166	172	179	185	192	199	206	212	219	226	233	239	246	253	260	267	274	281

α	n_2 / n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0.001		171	172	175	178	182	186	190	195	199	204	209	214	218	223	228	233	238	243	248
0.005		171	174	178	183	188	193	198	203	209	214	219	225	230	236	242	247	253	259	264
0.010	**18**	172	176	181	186	191	196	202	208	213	219	225	231	237	242	248	254	260	266	272
0.025		174	179	184	190	196	202	208	214	220	227	233	239	246	252	258	265	271	278	284
0.050		176	181	188	194	200	207	213	220	227	233	240	247	254	260	267	274	281	288	295
0.100		178	185	192	199	206	213	220	227	234	241	249	256	263	270	278	285	292	300	307
0.001		190	191	194	198	202	206	211	216	220	225	231	236	241	246	251	257	262	268	273
0.005		191	194	198	203	208	213	219	224	230	236	242	248	254	260	265	272	278	284	290
0.010	**19**	192	195	200	206	211	217	223	229	235	241	247	254	260	266	273	279	285	292	298
0.025		193	198	204	210	216	223	229	236	243	249	256	263	269	276	283	290	297	304	310
0.050		195	208	214	221	228	235	242	249	256	263	271	278	285	292	300	307	314	321	321
0.100		198	205	212	219	227	234	242	249	257	264	272	280	288	295	303	311	319	326	334
0.001		210	211	214	218	223	227	232	237	243	248	253	259	265	270	276	281	287	293	299
0.005		211	214	219	224	229	235	241	247	253	259	265	271	278	284	290	297	303	310	316
0.010	**20**	212	216	221	227	233	239	245	251	258	264	271	278	284	291	298	304	311	318	325
0.025		213	219	225	231	238	245	251	259	266	273	280	287	294	301	309	316	323	330	338
0.050		215	222	229	236	243	250	258	265	273	280	288	295	303	311	318	326	334	341	349
0.100		218	226	233	241	249	257	265	273	281	289	297	305	313	321	330	338	346	354	362

* Table entries are w_α of the Mann-Whitney test statistic T, where P(T< w_α) less than or equal to α.

If n_1 or n_2 is > 20, then approximate w_α by:

$w_\alpha = n_1(n+1)/2 + Z_\alpha\sqrt{n_1 n_2 (n+1)/12}$, where Z_α is the standard normal quartile and n = n_1 + n_2

Environmental Technology Verification

General Verification Protocol

Appendix C

Selected References

February 2007
(Rev. May 2013)

Environment Canada Environnement Canada

Environmental Technology Verification
Selected References

Analytical Procedures

Air

US EPA, "Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air", United States Environmental Protection Agency. Center for Environmental Research and Development. Cincinnati, OH, Report No. EPA/625/R-96/010b.(1999)

Water

Adams, V.D., "Water and Wastewater Examination Manual", Lewis Publishers, Chelsea, Michigan (1990).

APHA, AWWA, WPCF, "Standard Methods for the Examination of Water and Wastewater" 20th edition (1998).

Soil & Sediments

APHA, AWWA, WPCF, "Standard Methods for the Examination of Water and Wastewater" 20th edition (1998).

American Society of Agronomy, “Methods of Soil Analysis”, Madison, Wisconsin (1982).

Canadian Society of Soil Science, “Soil Sampling and Methods of Analysis”, Lewis Publishers (1993).

US EPA SW-846 “Test Methods for Evaluating Solid Waste, Physical/Chemical Methods”. URL: http://www.epa.gov/sw-846/sw846.htm

Experimental Design

Box, Hunter & Hunter, “Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building”, John Wiley & Sons, New York (1978).

Johnson, N.L. and Leone, F.C., “Statistics and Experimental Design in Engineering and the Physical Sciences” Vol. I, John Wiley & Sons, Inc., New York (1964).

Snedecor, G.W. and Cochran, W.G., “Statistical Methods” 6th Ed. Iowa State Univ. Press, Ames, Iowa (1971).

General

ASCE, "A Standard for the Measurement of Oxygen Transfer in Clean Water". American Society of Civil Engineers, New York, N.Y. (1984).

ASTM, "1996 Annual Book of ASTM Standards Section 11 Water and Environmental Technology" Volume 11.01 Water (I) and Volume 11.02 Water (II), American Society for Testing and Materials (1996).

ASTM, "Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method", ASTM Designation: E 691 - 05, ASTM Committee E-11, American Society for Testing and Materials .

CSA, "Guideline on Environmental Labeling" CSA Standard Z761-93, Canadian Standards Association (1993).

Laboratory Accreditation

CAEAL, the Directory of Certified Laboratories on CAEAL's web site at http://www.caeal.ca.

CSA, "Requirements for the Competence of Environmental Laboratories" CAN/CSA-Z753-95, Canadian Standards Association (1995).

Quality Assurance

Adams, V.D., "Water and Wastewater Examination Manual", Lewis Publishers, Chelsea, Michigan (1990).

APHA, AWWA, WPCF, "Standard Methods for the Examination of Water and Wastewater" 20th edition (1998).

ASTM, "Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method", ASTM Designation: E 691 - 92, ASTM Committee E-11, American Society for Testing and Materials (1992).

Foroutan, M., "Quality Control in the Wastewater Technology Centre Laboratory Division", Water Technology International Corporation, Environmental Chemistry Laboratory Division, (1993).

US EPA, "Preparing Perfect Project Plans - A Pocket Guide for the Preparation of Quality Assurance Project Plans", United States Environmental Protection Agency Risk Reduction Laboratory, Report No. EPA/600/9-89/087 (1989).

Sample Collection and Preservation

APHA, AWWA, WPCF, "Standard Methods for the Examination of Water and Wastewater" 20th edition (1998).

CCME, "Guidance Manual on Sampling, Analysis and Data Management for Contaminated Sites" Volume I: Main Report, CCME EPC-NCS62E, and Volume II: Analytical Method Summaries, CCME EPC-NCS66E, the National Contaminates Sites Remediation Program, Canadian Council of Ministers of the Environment (1993).

Foroutan, M., "Control Charts", Water Technology International Corporation, Environmental Chemistry Laboratory Division (1995).

Keith, L.H., "Environmental Sampling: A Summary", Environmental Science and Technology, Vol. 24, No. 5 (1990).

MOE - MISA, "Protocol for the Sampling and Analysis of Industrial/Municipal Wastewater", Municipal/Industrial Strategy for Abatement (MISA), Ontario Ministry of Environment & Energy, (199).

US DoI, "Recommended Methods for Water-Data Acquisition", Preliminary report of the Federal Interagency Work Group on Designation of Standards for Water Data Acquisition, U.S.

Department of the Interior, Geological Survey, Office of Water Coordination, Washington, D.C., (1972).

Statistical Data Analysis

Box, Hunter & Hunter, "Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building", John Wiley & Sons, New York (1978).

Johnson, N.L. and Leone, F.C., "Statistics and Experimental Design in Engineering and the Physical Sciences" Vol. I, John Wiley & Sons, Inc., New York (1964).

McClave, T., James. T., and Sincich T. Statistics. 9th edition. Prentice-Hall Inc. New Jersey ISBN 0 – 13- 065598 – 8 (2003).

Snedecor, G.W. and Cochran, W.G., "Statistical Methods" 6th Ed. Iowa State Univ. Press, Ames, Iowa (1971).

Steel, R.G. and J.H. Torrie, Priciples and Procedures of Statistics. McGraw-Hill Book Company, New York (1980).

US EPA, "Preparing Perfect Project Plans - A Pocket Guide for the Preparation of Quality Assurance Project Plans", United States Environmental Protection Agency Risk Reduction Laboratory, Report No. EPA/600/9-89/087 (1989).

World Wide Web Links

CAEAL	http://www.caeal.ca/
SCC	http://www.scc.ca
CSA	http://www.csa.ca
MOE	http://www.ene.gov.on.ca/
US EPA	http://www.epa.gov
EC	http://www.ec.gc.ca
Canadian ETV Program	http://www.etvcanada.ca
OCETA	http://www.oceta.on.ca

Environmental Technology Verification

General Verification Protocol

Appendix D

Glossary

February 2007
(Rev. May 2013)

Environment Canada Environnement Canada

Glossary of Terms for Environmental Technology Verification

The statistical definitions in this glossary are intended to provide a sufficient depth of understanding for the non-statistician to employ the ETV SAWs. The definitions are couched in the terminology and phrasing of the ETV program. Some of the definitions provided below may not be strictly correct in a mathematical sense.

Accreditation is recognition by an established (registered) organization for competence in performing high quality activities.

Alpha (α) See Type I error.

Alternative hypothesis the hypothesis that the technology developer wishes to accept or verify. Also, see null hypothesis.

Applicant (vendor) is the agent, supplier or manufacturer, who submits an environmental technology for verification through the ETV Program.

Assessment involves a review of data and information that describes the performance and integrity of an environmental technology or process. The effort applied to the assessment and the severity of the specifications in an assessment protocol dictate the merit of the result. In increasing order of rigor, the recognized products are: (1) peer review of data, (2) verification, (3) certification, (4) guarantee.

Audit involves a review of the performance and integrity of an environmental technology.

Average See mean.

Beta (β) See Type II error.

Canadian ETV Program is responsible for the management and delivery of the ETV Program and oversees each verification and issues the verification certificates. The Canadian ETV Program uses approved Verification Entities to conduct independent assessments of the technology performance claims.

CAEAL Canadian Association of Environmental Analytical Laboratories Inc., a partner in the SCC Program for the Accreditation of Laboratories (Canada), is an agency for certifying laboratories for the analyses of specific parameters. Some 87 labs have been certified under the Laboratory Certification Program. The most recent Directory is available at http://www.caeal.ca

Certification involves the repeated or ongoing assessment of a technology performance by an independent third party, based on it meeting some established set of standards(s). Certification also normally includes liability.

Chain-of-Custody refers to the ability to trace the possession and handling of the sample from the time of collection through analysis and final disposition, to ensure sample integrity from collection to data reporting (Standard Methods, 1992).

Confidence interval of α % for a parameter a range of values such that the true but unknown population parameter will fall within that range 100(1.00 – α) % of the time. Alternatively, the range of values, which upon repeated sampling, will enclose the true but unknown population parameter 100(1.00 – α) % of the time.

Conformance is an affirmation or judgment that an activity, product or process has met the requirements of the relevant specifications or standards.

Confounding factor any variable that is not of direct interest but may confound the interpretation of the verification experiment.

Counts **d**ata that may be counted. An example of count data is the number of fish in a tank. Count data takes on integer values.

CSA Canadian Standards Association.

Data set is a series of recorded observations that are specific to a single control or operating parameter or waste, feed or discharge characteristic.

Denominator the lower portion of a fraction.

Distribution this term is commonly used as a substitute for the phrase "probability distribution." A probability distribution describes how probable it is for an observation, or range of observations to occur for a given distribution. For example the probability distribution for the outcome of a dice toss is uniform. All tosses are equally likely.

Environmental Benefit is any significant alleviation of the detrimental effects that the creation, use or disposal of goods or services has on the environment and the health and welfare of humans and ecology.

Environmental technology (product or process) is a system consisting of equipment and/or materials, the operating procedures and the skills and knowledge to fulfill specified requirements for environmental performance, reliability and safety. Included should be the associated quality control elements that apply to the manufacturer and to the user of the technology.

Equipment-based environmental service is a service that can make claims solely on measurable performance of the equipment or technology under specified conditions.

Experiment any process run or evaluation that generates data suitable for testing a performance claim.

Experimental design the totality of knowledge used to design a verification experiment that addresses the performance claim in a cost-effective, legally defensible manner.

Experimental unit the smallest unit to which an experimental treatment may be applied. For example the experimental unit when considering a performance claim involving a valve modification to the carburetor of an internal combustion engine, is the carburetor. In a tank farm, the experimental unit would be a tank. Also see replicate.

Hypothesis testing is a statistical technique used to select one conclusion from two possible choices (null hypothesis, alternative hypothesis). This method is used when a decision requires a high degree of confidence, such as in the verification of a performance claim. The **Null Hypothesis (H_o)** is a simple statement that a statistic is mathematically acceptable or correct, when the hypothesis is evaluated within an acceptable range of probability (eg. at the 95% confidence level.) If the null hypothesis is concluded to be false, then the **Alternative Hypothesis (H_a)** will be assumed to be true.

Independent assessment is one by an individual (or organization) who does not gain benefits from the company being audited nor is influenced by payments or other benefits from the vendor and/or who is bound by a recognized professional code of ethics to report accurate information.

Integer a whole number greater than or equal to zero.

ISO International Organization for Standardization

Lead Organization is an independent company, which is responsible for the ETV program delivery. It oversees each verification project and issues the Verification Certificates. The Lead Organization uses Verification Agencies to conduct independent assessments of the technologies.

Mean a measure of the center of a data set. It is estimated by the sum of the measurements divided by the total number of measurements constituting the sum. The mean of n observations from the random variable x is given by:

$$\text{mean} = \sum_{i=1}^{n} x_i \Big/ n$$

Median the median is that observation which divides the ranked observations in half. For an even number of observations the median is the average of the two middle observations. The median is equivalently known as the 50^{th} percentile.

Normal distribution a specific probability distribution that is bell-shaped or symmetrical about a mean. The normal distribution has numerous attractive features that allow it to be widely used.

NSC National Standard of Canada

NSS National Standards System

Null hypothesis the hypothesis that the technology developer wishes to refute based upon the results of the verification experiment. For example a technology developer may claim that a technology produces a mean change of 20 units. The null hypothesis would be that the technology does not produce a mean change. Also see alternative hypothesis.

Numerator the upper portion of a fraction.

Observation A single data point such as a measurement, reading, etc. For a random variable y, the i^{th} observation is designated as y_i. Also see variable.

Peer Review Quality Information refers to technical materials accepted for publication in a refereed technical journal.

Performance claim is a measurable, reproducible, verifiable, and specific technology result that describes the performance of the environmental technology.

percentile Usually stated in association with a percentage. The observation below which, the stated percentage of observations lie. For example the 50^{th} percentile is the observation below which 50% of the ranked observations occur.

population The group of individuals, objects or items we wish to make inferences about. A population is the set of all elements, usually measured or derived characteristics or attributes that are of interest in a particular study.

power The probability of accepting the alternative hypothesis when it should be accepted. Mathematically this is equal to $1 - \beta$.

probability distribution See distribution.

proportion A fraction, ranging from, and including 0 to 1. The proportion is often multiplied by 100% and stated as a percentage.

Quality Assurance (QA) refers to a definitive plan for laboratory operation that specifies the measures used to produce data of known precision and reproducibility. Quality control (QC) refers to a set of measures within a sample analysis methodology to assure that the process is in control (Standard Methods, 1998). (include ISO Guide 25 info on Quality requirements)

Quantile usually stated in association with a proportion. The observation below which, the stated proportion of observations lie. For example the 0.5 quantile is the observation below which ½ of the ranked observations occur.

Random variable a characteristic or attribute that exhibits variability. For example the weight of a person is a random variable.

Real number a real number is the set of all possible infinite decimal expansions. Mathematically it is the union of the rational and irrational sets.

Reference Laboratory is a laboratory operated for the principal purpose of analyzing samples referred to by other laboratories for confirmatory analysis. A reference laboratory conducts quality assurance functions relative to other laboratories and may perform unusual, highly specialized, and difficult analyses not generally available through commercial laboratories.

Replicate measurements of a variable from different experimental units. When applying treatments to tanks in a tank farm, each tank is an experimental unit. Repeated measurements from a single tank are sub samples or repeated measures, not replicates.

Robust Procedures are statistical procedures that are insensitive to deviations from normality in the data.

Run see Sample campaign.

Sample the subset of the population collected, from which inferences will be made. When performing statistical tests we are always working with a sample.

Sample campaign (Run, Test) represents the complete number of data sets that are required to assess the performance of an environmental technology, including the control parameters and the feed and discharge characteristics.

Sample population is a representative data set for a particular parameter.

SCC Standards Council of Canada

Small sample When making inferences from normally distributed variables with sample sizes less than 30, quantiles and probability values should be obtained from Student's t-distribution.

Standard deviation is a measure of the spread of the data defined as the square root of the variance.

Statistical distribution See distribution.

Symmetric usually used in conjunction with the phrase "probability distribution". A probability distribution is symmetric if the mean is approximately equal to the median and the probability that a random variable, X is less than any percentile, $p_i < 0.5$ is equal to the probability that X is greater than the 1- p_i^{th} percentile.

Test see Sample campaign.

Testing Agency is an organization that has been contracted by the technology vendor or Verification Entity (at the technology vendor's request) to evaluate a technology and

collect additional data, following appropriate ETV Test Protocols to provide adequate information to permit completion of a performance claim verification. Where appropriate, accredited testing laboratories should be used.

Test Protocol is the detailed procedures for generating data that qualify a technology for verification.

Treatment an environmental technology (procedure/process/additive/modification) whose effect will be compared with another treatment such as: the absence of the environmental technology, a standard technology, alternative technologies, etc.

Type I error the probability of rejecting the null hypothesis when it should be accepted. This is set to 0.05 or 5% for the ETV program. The type I error is conventionally designated as α**.**

Type II error the probability of accepting the null hypothesis when it should be rejected. The type II error is conventionally designated as β.

Validation is the confirmation, by the examination and provision of objective evidence, that the particular requirements for a specified application are fulfilled.

Variable a characteristic that exhibits variability. For example, the BOD_5 of an effluent sample is a variable. A single measurement of the BOD_5 is an observation.

Variance a measure of the dispersion or spread of the data around the mean value. It is the sum of the squared deviations from the mean. The **population variance** may be estimated as:

$$\text{variance} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n},$$

where n is the sample size and x is the variable for which the variance is being estimated. The **sample variance** may be estimated as:

$$\text{variance} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right].$$

Vendor see Applicant

Verification is an examination of environmental performance claims made by suppliers, and of available supporting information, for the purpose of validating the performance claims. The purpose of verification is to substantiate that the performance and integrity of the environmental technology satisfies a standardized protocol as specified by Environment

Canada's ETV Program. The verification must include the confirmation, by examination and provision of objective evidence, that specified requirements are achieved. These specifications must include that an environmental product or process is based on sound scientific and engineering principles, that it is effective, reliable and protective of health and environment, and that it performs in this manner under defined operating and environmental conditions.

Verification Certificate is a single page document that includes the ETV "Seal of Approval" which acknowledges that the performance claim has been verified. The Verification Certificate will definitively detail the performance claim and identify all relevant accompanying documentation that validates the technology claim. Verification Certificates are valid for 3 years and may be withdrawn prior to that time.

Verification Entity is an approved (by the Canadian ETV Program) independent organization or technical expert that conducts environmental technology performance claim assessments and activities to validate the claim. The Verification Entity must have independence and objectivity in terms of potential financial, fiduciary, procedural and/or technical relationships and must have the technical capability and/or engineering expertise to perform its roles and responsibilities as a Verification Entity. The Verification Entity will also be expected to provide expert technical advice to the Canadian ETV Program. The Verification Entity will have gone through a screening process and will be approved for specific technologies and working languages. It is the Verification Entity's responsibility to notify the Canadian ETV Program of any change in verification staff or resources that could impact their status.

List of Commonly Used Symbols

Parameters	Population Estimate	Sample Estimate
mean	μ	$\bar{x}$
variance	σ^2	s^2

Others	Symbol
the probability of making a Type I error	α
the probability of making a Type II error	β
data point or observation	x
i^{th} data point or observation	x_i
degrees of freedom	v
F distribution with v_1, v_2 degrees of freedom	$F_{v1,v2}$
the capital Greek letter sigma denotes addition over the limits given.	$\sum_{i=1}^{n}$
sample size	n

List of Abbreviations

effective degrees of freedom	dof_e

The population estimate is an estimate of the set of all elements (measured or derived characteristics or attributes) that are of interest in a particular study.

The sample estimate is an estimate of a portion or subset of the population.

Environmental Technology Verification

General Verification Protocol

Appendix E

Case Study 1

February 2007
(Rev. May 2013)

Environment Canada Environnement Canada

Performance Claim Evaluation of the

ENVIRO-MASTER BIOFILTER

Case Study 1

Prepared for
Canadian ETV Program

By

Pollutech International Limited

Revised by Canadian ETV Program on:
September 2002
&
February 2007
(Rev. May 2013)

TABLE OF CONTENTS

1.0 INTRODUCTION 1

2.0 INFORMATION SUPPLIED BY APPLICANT 2

2.1 The Applicant's Technology Information 2
2.2 Technical Specifications 2
2.3 Performance Claim 2
2.4 Process Application Schematic 3
2.5 Sampling & Analytical Protocols 3
2.6 Data Sets 4
2.7 Literature Citations for Enviro-Master 6

3.0 REVIEW OF TECHNOLOGY CLAIM APPLICATION 7

3.1 Review of Application 7
3.2 Comments 9
3.3 Conclusion 10

4.0 Review of Technology 10

4.1 Test Conditions for Data to be Verified 13
4.2 Soundness of Technology 13
4.3 Health & Safety Issues 13
4.4 Training 14
4.5 Discussion 14
4.6 Conclusions 15

5.0 REVIEW OF DATA 16

5.1 Introduction 16
5.1.1 Comments for Table 7 20
5.2 Data Analysis 21

6.0 STATISTICAL EVALUATION OF CLAIM 22

6.1 Introduction 22
6.2 Statistical Analysis Worksheet No. 1 Assessing Normality of Data 23
6.3 Statistical Analysis Worksheet No. 5 Testing Mean is Equal to a Specified Value 27

7.0 CONCLUSIONS 31

7.1 Limitation of Verification 31
7.2 Lessons from Verification 31

Appendix A- ASSUMING DATA IS NOT NORMALLY DISTRIBUTED 33

Statistical Analysis Worksheet No. 8 Testing Median is Equal to a Specified Value: Effluent BOD_5 33
Statistical Analysis Worksheet No. 8 Testing Median is Equal to a Specified Value: Effluent TSS 36

LIST OF TABLES

Table 1: Performance Data Set # 1 4
Table 2: Performance Data Set # 2 5
Table 3: Split Sample Data Set # 3 6
Table 4: Application Review Checklist – Mandatory Information 7
Table 5: Technology Review 11
Table 6: Study Design Checklist 16
Table 7: Data Validity Checklist 19
Table 8: Performance Data Set # 1 22
Table 9: Spreadsheet for Normality Calculations - BOD_5 effluent data 24
Table 10 Spreadsheet for SAW # 5 - Calculation of Variance, Effluent BOD_5 29
Table 11 Results of Statistical Analyses 29
Table 12: Student's 't' Values 30
Table 13: Summary of Claim Evaluations 30

1.0 INTRODUCTION

The following is a hypothetical technology verification claim. It is a case study designed to illustrate information and data shortcomings and gaps. The aim of this case study is to lead the course participant through the steps, which need to be taken when verifying a technology performance claim made by the technology proponent. It is for instructional purposes only.

It is intended to remind the VE of the two criteria, which a technology must meet, in order to be ***eligible*** for the ETV Program:

1. It must be either:
    a) an environmental technology or process that offers an environmental benefit or addresses an environmental problem, or
    b) an equipment-based environmental service that can make claims based solely on measurable performance of the equipment.

2. The claim must be:
    - specific and unambiguous.
    - meaningful and nontrivial.
    - measurable and verifiable.

For a claim to be ***verified***, the following three basic criteria must also be fulfilled:

1. The technology is based on sound scientific and engineering principles.
2. The claim is fully supported by peer-review quality data, which are supplied by the applicant.
3. The conditions of performance for the claim are clearly defined.

As well, to be eligible for receipt of a Verification Certificate, the technology must be currently commercially available or commercially ready for full-scale application.

It is important that the Verification Entity (VE) keep ***all*** of these criteria firmly in mind when going through the various technology claim verification steps.

The case study is not intended to endorse a particular product, nor does it constitute a performance claim. The case studies are used to provide guidance when statistically testing performance claims in the context of the ETV program. The rationale behind subjective decisions regarding level of significance, definition of small sample sizes and choice of statistical test methods endorsed, are given in the introduction to the SAWs, Appendix A.

2.0 INFORMATION SUPPLIED BY APPLICANT

The applicant supplied the following information:

1. information regarding the technology
2. some technical specifications
3. the performance claim
4. a schematic of the process application
5. a description of the sampling and analytical protocols used
6. three (3) data sets supporting the performance claim, and
7. three literature citations for the Enviro-Master.

2.1 The Applicant's Technology Information

The Enviro-Master Biofilter is the solution for all septic system applications. A select combination of advanced treatment processes ensures effective, dependable operation with minimal maintenance for years to come. The Enviro-Master Biofilter is a pre-engineered system that is delivered complete and ready to install wherever environmental protection and peace of mind are required.

2.2 Technical Specifications

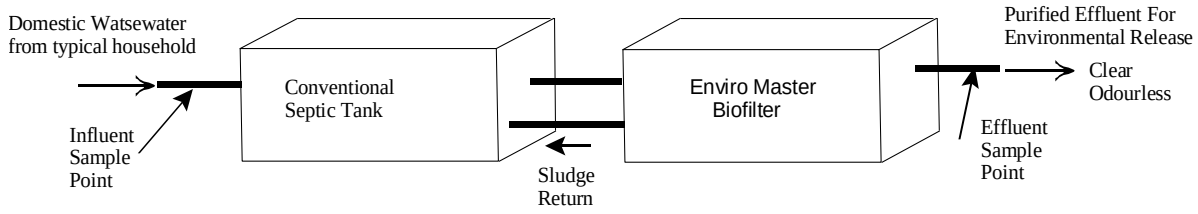
- advanced hydrophilic polymer growth substrate;
- interlocking, space-efficient, UV-resistant filtration media;
- quiescent clarification for effluent polishing; and,
- polycarbonate enclosure protects all internal components from the elements (optional colours available).

2.3 Performance Claim

The Enviro-Master Biofilter treats household wastewaters, to produce a purified effluent, which can be returned safely to the environment. Installed and operated by factory-trained and qualified technicians, the Enviro-Master Biofilter reduces bio-chemical oxygen (BOD) and total suspended solids (TSS) to minimal levels. The purified effluent can be released directly to the environment. No further treatment is required.

2.4 Process Application Schematic

Figure 1: Schematic of Enviro-Master Biofilter



2.5 Sampling & Analytical Protocols

Through 1996, grab samples of the influent and the effluent were taken monthly by a trained technician employed by Enviro-Master. The purpose of the sampling was to demonstrate the system's performance and to document the effluent quality for inclusion in the company's descriptive brochures. The samples were analyzed for total suspended solids (TSS) and 5-day biochemical oxygen demand BOD_5. The BOD samples were packed on ice and taken directly to the Blue Sky Analytical Services lab for immediate analysis. The BOD_5 and TSS analyses were done in accordance with Standard Methods (APHA) at Blue-Sky labs. The Blue-Sky lab is accredited by CAEAL. This data set is labeled as performance data set # 1.

Additional sampling was done in early 1997. The same technician visited the demonstration site daily to inspect the performance and to take samples of the influent and effluent. Samples were taken on 20 of the 35 sampling days for BOD5 analyses while samples for TSS analyses were taken daily. Again, Blue Sky Analytical Services lab performed the BOD_5 and TSS analyses that were done in accordance with Standard Methods (APHA). On two occasions the technician was accompanied to the site by an inspector from the County Health Unit. On both occasions, the sample taken by the technician was split; the inspector took half of each sample and submitted it to the Regional Health Unit labs for BOD_5 and TSS analysis. On three other occasions, the technician sent split samples directly to the Regional lab without the inspector being on site. This data set is labeled as performance data set # 2.

2.6 Data Sets

The following three data sets were provided with the application:

Table 1: Performance Data Set # 1

Sample Date	BOD_5 (mg/L)		TSS (mg/L)		Sample Type
	Influent	Effluent	Influent	Effluent	
17-Jan-96	153	16	196	19	Grab
16-Feb-96	242	15	162	20	Grab
17-Mar-96	247	17	179	17	Grab
17-Apr-96	243	18	183	24	Grab
17-May-96	231	15	160	14	Grab
17-Jun-96	203	18	142	7	Grab
17-Jul-96	163	18	170	16	Grab
16-Aug-96	136	17	184	10	Grab
16-Sep-96	147	17	213	18	Grab
16-Oct-96	261	20	185	9	Grab
16-Nov-96	135	18	174	17	Grab
16 Dec-96	227	18	154	17	Grab

Table 2: Performance Data Set # 2

Sample Date	BOD_5 (mg/L)		TSS (mg/L)		Sample Type
	Influent	Effluent	Influent	Effluent	
01-Jan-97	295	20	94	24	Grab
02-Jan-97	246	18	136	28	Grab
03-Jan-97	148	17	195	23	Grab
04-Jan-97	266	16	196	16	Grab/Split
05-Jan-97	135		116	9	Grab
06-Jan-97			194	4	Grab
07-Jan-97	280	17	157	21	Grab/Split
08-Jan-97	254	13	255	16	Grab
09-Jan-97	252	17	149	21	Grab
11-Jan-97			203	24	Grab
11-Jan-97			135	17	Grab
12-Jan-97	176	18	101	12	Grab
13-Jan-97	282	16	162	22	Grab
14-Jan-97			177	21	Grab
15-Jan-97			185	32	Grab
16-Jan-97	217	13	188	21	Grab/Split
17-Jan-97	236	18	179	17	Grab
18-Jan-97			178	8	Grab
19-Jan-97	199	16	165	14	Grab
20-Jan-97		14	221	19	Grab
21-Jan-97	221		131	22	Grab
22-Jan-97	106	19	142	12	Grab
23-Jan-97			159	10	Grab
24-Jan-97	157	17	193	7	Grab/Split
25-Jan-97			182	22	Grab
26-Jan-97			174	16	Grab
27-Jan-97	180	18	228	28	Grab
28-Jan-97	294	18	124	20	Grab
29-Jan-97	311	18	209	22	Grab
30-Jan-97			270	16	Grab
31-Jan-97	216	15	139	9	Grab/Split
01-Feb-97			214	25	Grab
02-Feb-97			229	8	Grab
03-Feb-97			266	19	Grab
04-Feb-97			271	17	Grab

Table 3: Split Sample Data Set # 3

Sample Date	BOD_5 (mg/L)		TSS (mg/L)		Sample Type
	Influent	Effluent	Influent	Effluent	
04-Jan-97	260	17			Grab/Split
07-Jan-97	280	17	16	8	Grab/Split
16-Jan-97	220	12			Grab/Split
24-Jan-97	150	15	15	5	Grab/Split
31-Jan-97	210	16			Grab/Split

2.7 Literature Citations for Enviro-Master

Tode, N. H. and S. Q. Loam. 1995. Ground disposal of treated effluent. Subsurface Digest, 21:333-342.

Vadose, Z. and E. Vapo. 1991. On-site technologies for wastewater treatment. Wastewater Weekly. 65:79-87.

Enviro-Master Technologies Inc. Worry-free treatment for all applications. Technical Specification Sheet # 97-A01

Armed with this information we are now ready to start our evaluation procedure.

3.0 REVIEW OF TECHNOLOGY CLAIM APPLICATION

This Section provides a summary of the information provided by the applicant as included with a pre-screening application form and the formal application form submitted to the Canadian ETV Program and reviewed by the VE for the ETV program. If information vital to the verification process is missing or if clarification for the VE is required, this can be requested by the VE and supplied by the applicant through verbal/written communications with the applicant while the verification is in progress.

3.1 Review of Application

The technology and all information provided by the Applicant with the Formal Application and all subsequent transmittals to the Verification Entity are now reviewed and summarized using the Checklist (Table 4). Their purpose is to assist you in deciding how to answer the questions, either YES or NO and comments as to how the criteria is applicable to this case.

The criteria ratings have been left blank on purpose. It is left to us to fill them in. Let's spend some time going through Table 4 and rate each item.

Table 4: Application Review Checklist – Mandatory Information

Ref.	Criteria		Information Provided	
			Yes[1]	No
1.1	Signed Formal Application		☐	☐
1.2	Signed Declaration Regarding Codes & Standards submitted with signed formal application		☐	☐
1.3	Technology provides an environmental benefit.		☐	☐
1.4	A copy of "Claim to be Verified" for each performance claim to be verified included with the Formal Application.		☐	☐
1.5	Performance Claim composed in a way that satisfies "Criteria for Specifying Claims" :		☐	☐
	1.5.1	Include Technology name (and model number)	☐	☐
	1.5.2	Include application of the technology	☐	☐
	1.5.3	Include specific operating conditions during testing	☐	☐
	1.5.4	Does it meet minimum Canadian Standards/Guidelines *	☐	☐
	1.5.5	Does it specify the performance achievable by the technology	☐	☐
	1.5.6	Is it the performance measurable	☐	☐
1.6	Standard operating practices and a description of operating conditions for each individual performance claim specified.		☐	☐
1.7	The proponent has supplied significant references describing or supporting scientific and engineering principles of the technology. (see Chapter 4)		☐	☐
1.8	Two or more names and contact information of independent (no vested interest in the technology) experts, qualified (backgrounds of experts are needed) to discuss scientific		☐	☐

[1] Provide written justification for yes or no information provided.

Ref.	Criteria	Information Provided	
		Yes[1]	No
	and engineering principles on which the technology is based. These experts must be willing to be contacted by the VE.		
1.9	Brief summary of significant human or environmental health and safety issues associated with the technology. (Note: this criterion complements but does not replace the obligation for the applicant to submit a duly signed "Declaration Regarding Codes and Standards")	☐	☐
1.10	Brief summary of training requirements needed for safe, effective operation of technology, and a list of available documents describing these requirements. (Note: this criterion complements but does not replace the obligation for the applicant to submit a duly signed "Declaration Regarding codes and standards")	☐	☐
1.11	Process flow diagram(s), design drawings, photographs, equipment specification sheets (including response parameters and operating conditions), and/or other information identifying the unit processes or specific operating steps in the technology. If feasible, a site visit to inspect the process should be part of the technology assessment.	☐	☐
1.12	**Supplemental materials** (optional) have been supplied which offer additional insight into the technology application integrity and performance, including one or more of :		
	A copy of patent(s) for the technology, patent pending or submitted.	☐	☐
	User manual(s).	☐	☐
	Maintenance manuals.	☐	☐
	Operator manuals.	☐	☐
	Quality assurance procedures.	☐	☐
	Sensor/monitor calibration program.	☐	☐
	Certification for ISO 9001, ISO 14000, or similar program.	☐	☐
	Material Safety Data Sheet (MSDS) information.	☐	☐
	Workplace Hazardous Materials Information System (WHMIS) information.	☐	☐
	Health and Safety plan.	☐	☐
	Emergency response plan.	☐	☐
	Protective equipment identified.	☐	☐
	Technical brochures.	☐	☐
1.13	The applicant provided adequate documentation and data. There is sufficient information on the technology and performance claim for the performance claim verification. [If necessary, the VE should communicate with the Canadian ETV Program to request copies of the necessary documentation and required data that are available to support the claims.]	☐	☐

3.2 Comments

Now it would be in order to comment regarding patents application/date of receipt and whether documentation was provided for any equipment specifications, process flow diagrams, user manuals, maintenance procedures, Material Safety Data Sheet (MSDS) or Workplace Hazardous Materials Information Sheet (WHMIS), ISO Certification or emergency plans.

We now make a professional judgment as to whether the applicant supplied adequate documentation to satisfy the various requirements outlined in Table 4. If the answer is in the affirmative, then we proceed to the next step, a review of the Enviro-Master Biofilter technology.

There are several things we note: Lets look at the claim made.

"The Enviro-Master Biofilter treats household wastewaters, producing a purified effluent which can be returned safely to the environment. Installed and operated by factory-trained and qualified technicians. The Enviro-Master Biofilter reduces bio-chemical oxygen demand (BOD) and total suspended solids to minimal levels. The purified effluent can be released directly to the environment. No further treatment is needed."

This performance claim,

- does not explicitly state the hypotheses being tested
- does not qualify standard operating conditions, or ranges of influent
- does not make specific performance claims, and
- contains ambiguous phrasing such as the phrase "purified effluent".

The performance claim must be restated so that it is:

- specific and unambiguous,
- meaningful and nontrivial as well as measurable and verifiable.

Once we have done this, stating the claim as a series of hypotheses, then we can proceed with the evaluation of these hypotheses objectively using an appropriate statistical methodology.

When reviewing the information submitted by the applicant, we notice the following:

- The split sample TSS data in Table 3 is really useless. Something went wrong in the analysis of the influent TSS. The analyst should have been alerted by the fact that as a rough rule of thumb, the TSS influent values should be about 20% less than the BOD_5 influent values. This is attributable to the settling of the raw sewage in a septic tank (Figure 1). So we conclude that quality control was absent.

- The performance data as noted in Table 2 only covered the month of January (winter). In the absence of any wastewater temperature data we are reluctant to combine the data for January with those of Table 1 which were collected on a monthly basis, year-round. However if we can satisfy ourselves that there is really no difference between the BOD_5 mean of Table 1 and that of Table 2, then we can combine the data. The same would have to hold for the TSS data.

- We could also elect to test our hypothesis using each data set separately. The winter data (Table 2) would show us whether or not the claim made is robust for winter operation. This is important as we are evaluating a claim for a biological system, and temperature plays an important role. What would have been good, are influent and effluent temperatures for the Enviro-Master Biofilter. This illustrates the importance of good planning of the type of data that should be collected.

- If we combine the two data sets, #1 and #2, then by increasing the sample size, we increase the power of the statistical test. So this is highly desirable. In the final analysis we must use our professional judgment in whether or not we should pool the data sets.

These points serve to illustrate that performance claim evaluations cannot be approached in a regimented manner.

Remember that statistical tests are only tools. They are not a substitute for best professional judgment.

3.3 Conclusion

The performance claim should be restated as:

1. **The Enviro-Master Biofilter reduces BOD_5 to at least 20 mg/L under the specified operating conditions (i.e. influent BOD_5 < 200mg/l).**

2. **The Enviro-Master Biofilter reduces TSS to at least 25 mg/L under the specified operating conditions (i.e. influent TSS < 200 mg/L).**

Let's assume that we are satisfied with the information supplied with the application. We now proceed with the technology review.

4.0 Review of Technology

We now go back and look at the description of the technology, and get a firm picture of what the technology does and how this is accomplished. We use the questions noted in Table 5 to help us do this. The questions deal with the description of the technology, any applicable environmental standards and whether or not the process is commercially ready. We also note whether or not the applicant has identified process operating parameters and whether the effects of variable input/output parameters are understood by the technology proponent.

The VE is also required to supply the rationale applied in certain judgments as shown in Table 5. These would be summarized in a Section that would typically follow Table 5. Their purpose is to assist you in deciding how to answer the questions, either YES or NO with any pertinent explanations.

Now let's go through Table 5 and see how each item should be rated.

Table 5: Technology Review Criteria Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
Technology Description			
2.1	Technology based on scientific and technical principles. (It will be necessary for the VE to read the key articles and citations listed in the Formal Application. It may also be necessary to contact the independent experts listed in the Formal Application to obtain additional information.)	☐	☐
2.2	Technology supported by peer review technical literature or references. (Peer review literature and texts must be supplied with the Formal Application as well as relevant regulations and standards that are pertinent to the performance claim)	☐	☐
2.3	Technology designed, manufactured, and/or operated reliably. (historical data from the applicant, not conforming to all data criteria, may be useful for the VE to review to assess the viability of the technology not for verification, but for insight purposes)[3]	☐	☐
2.4	Technology designed to provide an environmental benefit and not create an alternative environmental issue. (e.g. it does not create a more hazardous and or unmanaged byproduct and it does not result in the transfer of an environmental problem from one media to another media without appropriate management of the subsequent contaminated media)	☐	☐
2.5	Technology conforms to standards for health and safety of workers and the public.[4] The vendor must submit a signed "Declaration Regarding Codes & Standards", with the Formal Application. The role of the Verification Entity is to ensure this signed document is included with the information that is reviewed for the performance claim verification	☐	☐

[2] Provide written justification for no meets criteria.

[3] *Also note* ***The VE should use best judgment and apply standards relevant to the technology sector to generally assess whether the technology has been designed and manufactured in an acceptable fashion. A critical assessment of the materials / apparatus used in the technology is beyond the scope of the ETV program. Any assessment of the integrity of the manufacture of technology components must be performed by personnel whose experience and expertise qualify them to undertake this activity. It is not the responsibility of the Verification Entity to assess the integrity of materials and substances used in the manufacture of the technology, other than to understand their use and implication on the performance of the technology.***

It is the vendor's responsibility to ensure that applicable regulations and guidelines are satisfied with respect to application of the technology. The vendor must submit a signed "Declaration Regarding Codes & Standards", generally with the Formal Application. The role of the Verification Entity is to ensure this signed document is included with the information that is reviewed for the performance claim verification.

Claim verification by the Verification Entity does not represent any guarantee of the performance or safety of the equipment or process. The Verification Entity shall not be liable in any way in the event that the device or process fails to perform as advertised by the supplier or as expected by the consumer. The Verification Entity shall not be liable for any injury to person or property resulting from the use of the equipment or process.

[4] For the purposes of the ETV Program, the health and safety issue has been defined as a subjective criteria, requiring a value judgment on the part of the reviewer as to the integrity or reliability of any or all health and safety

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
Environmental Standards			
2.6	Technology achieves federal, provincial, and/or municipal regulations or guidelines for management of contaminated and or treated soils, sediments, sludges, or other solid-phase materials.	☐	☐
2.7	Technology achieves federal, provincial, and/or municipal regulations or guidelines for all (contaminated and or treated) aqueous discharges as determined by the applicants information.	☐	☐
2.8	Technology achieves federal, provincial, and/or municipal regulations or guidelines for all (direct or indirect) air emissions. If the environmental technology results in the transfer of contaminants directly or indirectly to the atmosphere, then, where required, all regulations or guidelines (at any level of government) relating to the management of air emissions must be satisfied by the applicant's information.	☐	☐
Commercial Readiness			
2.9	Technology and all components (apparatus, processes, products) is full-scale, commercially-available, **or alternatively see 2.10 or 2.11**, and, data supplied to the Verification Entity is from the use or demonstration of a commercial unit.	☐	☐

documentation provided by the applicant. As such, the Verification Entity cannot assume any liability in making a "Best Professional Judgment" assessment of the technology using these criteria.

(continued footnote 5 from Ref. 2.5) A critical validation of the Health and Safety aspects of the vendor's technology is beyond the scope of the ETV program. Any validation of health and safety issues must be performed by personnel whose experience and expertise qualify them to undertake these activities. Staff from noted organizations and agencies [e.g., Health and Welfare Canada (H&W), Provincial Labour Ministries, Industrial Accident Prevention Association (IAPA), [US] Occupational Safety and Health Association (OSHA), water pollution control agencies, province/state health departments, fire protection associations, etc.], may be able to provide advice or technical services on these issues. It is **NOT** the responsibility of the Verification Entity to validate the Health and Safety aspects of the technology.

It is the vendor's responsibility to ensure that regulations and guidelines are satisfied in the application of the technology. The Verification Entity can request additional written confirmation from the applicant that the company has sufficient documentation to address worker health and safety issues and requirements related to the use of the technology, including an Emergency Response Plan.

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
2.10	Technology is a final prototype design prior to manufacture or supply of commercial units, **or alternatively see 2.11**, Note: Verification of the performance claim for the technology is valid if based on a prototype unit, if that prototype is the final design and represents a pre-commercial unit. The verification will apply to any subsequent commercial unit that is based on the prototype unit design. The verification will not be valid for any commercial unit that includes any technology design change from the prototype unit used to generate the supporting data for the verification.	☐	☐
2.11	Technology is a pilot scale unit used to provide data which when used with demonstrated scale up factors, proves that the commercial unit satisfies the performance claim.[5]	☐	☐
Operating Conditions			

4.1 Test Conditions for Data to be Verified

Here we make a judgment, based on the information at hand, which allows us to confirm that the testing and analytical protocols used, conformed to the rigor acceptable to scientific and engineering investigations of this type.

4.2 Soundness of Technology

To give us a comfort feeling about the technology, we judge if the technology is considered to be sound according to scientific and engineering principles used and/or established in textbooks, peer-reviewed journal articles, technical documents and/or patents. As well, data related to the technology must be of sufficient quality to pass a review by other technical experts.

Two other issues have to be addressed:

1. Health & Safety, and
2. Training

4.3 Health & Safety Issues

A critical validation of the Health and Safety aspects of the applicant's technology is beyond the scope of the ETV program. Personnel whose experience and expertise qualify them to undertake this activity must perform any validation of Health and Safety issues. Staff in organizations and agencies, such as Health and Welfare Canada (H&W), Provincial Ministries of Labour (MOL), Industrial Accident Prevention Association (IAPA), [US] Occupational Safety and Health Association (OSHA) can provide advice or technical services on these issues. It is

[5] In exceptional situations, data from a pilot scale unit may be used to validate a performance claim. This situation can be permitted if the pilot scale unit is a "scaled down" model of a full size commercial unit and engineering scale-up factors have been provided by the applicant as part of the verification process. The performance claim verification must include validating the scale-up factors.

NOT the responsibility of the Verification Entity to validate the Health and Safety aspects of the technology. It is the applicant's responsibility to ensure that such codes, regulations and guidelines are satisfied in the application of the technology.

The role of the Verification Entity is simply to request written confirmation from the vendor that the company has sufficient documentation to address worker health and safety issues and requirements related to the use of the technology. The Verification Entity is not responsible for reviewing or commenting on the appropriateness and sufficiency of this documentation.

A copy of the "**Declaration Regarding Codes & Standards**", signed by the applicant should be provided.

If however, the VE feels that the technology may pose a risk to safety, health or the environment, the VE stops the verification process and advises the Canadian ETV Program of these concerns. The VE does not restart the verification process until the applicant has addressed the VE's concerns to their satisfaction.

4.4 Training

The training issues associated with this technology were not specifically addressed in the documents received by the VE.

4.5 Discussion

Based on all the information received, we judge that the documentation and data provided by the applicant is adequate for the verification to proceed by applying the General Verification Protocol.

The applicant provided 3 articles written by their scientific staff. Two of the articles were published in journals that are not peer-reviewed. The other publication was a company article.

Initial documentation was provided and reviewed with the applicant. Subsequent to discussions with the applicant, the VE prepared a list of detailed questions concerning the process, data, and analyses. These were forwarded to the applicant. Responses were received that adequately describe the data collection procedures used and some of the process operating issues raised. A number of review meetings and follow up responses provided the VE experts with a degree of confidence in the process operation and results tabled by the applicant.

Major documents reviewed by the VE experts included Manufacturing technical reports, an outline of production protocol, photos of process operations, journal articles, sampling and analysis protocols and QA/QC documents.

4.6 Conclusions

We are now ready to state our conclusions concerning the verification done to this point. We are satisfied that:

1. **adequate information was provided to satisfy the conditions outlined in Table 4, Application Review.**

2. **the Enviro-Master Biofilter technology meets the ETV protocol requirements outlined in Table 5, Technology Review.**

3. **the Enviro-Master technology is sound.**

4. **process operating conditions were what are to be expected under normal process application.**

Health & Safety Issues, while beyond the scope of this verification process, were noted. The VE feels that the technology does not pose a risk to safety, health or the environment, if operated according to the according to the manufacturers instructions.

Training, was not specifically addressed, but some documentation was received which leads the Verification Entity to believe that the appropriate level of training will be implemented by Enviro-Master. This is reinforced by the knowledge that the Enviro-Master Biofilter technology is a commercial product and its success in the market place is very much dependent on producing a consistent quality product.

We are now ready to proceed to the next step, the assessment of the data quality.

5.0 REVIEW OF DATA

5.1 Introduction

We now have to judge the acceptability of the data generation process used by the applicant. The purpose of this phase of our evaluation process is to ascertain, on a qualitative basis, whether there was a specific or any experimental design approach used in deciding what variables should be measured and whether data collection was conducted with rigor and followed accepted scientific protocols. Only if the answer is in the affirmative, will the actual data underlying the claim made, be subjected to a variety of statistical procedures so as to verify the claim(s) made.

We divide this data review into the following steps:

1. an assessment concerning the design of the study, and
2. an assessment concerning the validity of the data.

To assist us in this task we have 2 checklists, as shown in Tables 6 and 7.

Table 6: Study Design Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[6]
3.1	Was a statistician, or an expert with specialized capabilities in the design of experiments, consulted prior to the completion of the test program, and if so please provide the contact details.[7]	☐	☐
3.2	Is a statistically testable hypothesis or hypotheses provided? (so that an objective, specific test is possible)[8]	☐	☐
3.3a-c	Does the verification study generate data suitable for testing the hypothesis being postulated? [9] Namely:	☐	☐
3.3a	Does the study measure the parameters used in the performance claim hypothesis?	☐	☐
3.3b	Does the study control for extraneous variability?	☐	☐
3.3c	Does the study include only those effects attributable to the environmental technology being evaluated?	☐	☐
3.4	Does the verification study generate data suitable for analysis using the generic SAWs? (i.e. it is clearly preferable that tests are designed with the SAWS in mind before test plans are written)	☐	☐
3.5	Does the verification study generate data suitable for analysis using other generic experimental designs (ANOVA etc)? (clearly, verification studies should be designed with the final data analysis in mind to facilitate interpretation and reduce costs)	☐	☐
3.6	Are the appropriate parameters, specific to the technology and performance claim, measured? (it is essential that the VE and the technology developer ensure that all parameters – e.g. temperature etc - that could affect the performance evaluation are either restricted to pre-specified operating conditions or are measured)	☐	☐
3.7a-d	Are samples representative of process characteristics at specified locations?. **namely:**	☐	☐
3.7a	Are samples collected in a manner that they are representative of typical process characteristics at the sampling locations for example the samples are collected from the source stream fully mixed etc	☐	☐

[6] Provide written justification for yes or no meets criteria.
[7] An expert statistician can help determine during the experimental design which experimental variables need to be controlled and or monitored so as to be able to defend a verification claim
[8] The hypothesis that Statistical Analysis Worksheets will test are of the general form:
What is the degree of confidence around a measured result?
Is a mean equal to a specified value?
Is a median equal to a specific value?
Is mean 1 = mean 2 ?
Is median 1 = median 2 ?
Is variance 1 = variance 2 ?
Can a process change an influent/product/waste by 'p' percent?
Are two paired measurements different?
[9] Note: When data are not available on a specific parameter, it may be possible to use data on a surrogate parameter that has known correlation to the unmeasured parameter. In this case, the correlation must be clearly defined, demonstrated and based on sound scientific, engineering and or mathematical principles. The applicant must submit that data for their set of tests.

3.7b	Is data representative of the current technology?	☐	☐
3.7c	Have samples been collected after a sufficient period of time for the process to stabilize?	☐	☐
3.7d	Have samples been collected over a sufficient period of time to ensure that the samples are representative of process performance?	☐	☐
3.8	Are samples representative of operating conditions? Note: A time lag occurs between establishing steady state conditions and stabilization of the observed process performance. This time lag depends in part on the time scale of the process. (i.e. for a Completely Stirred Tank Reactor (CSTR) flow-through system, the time scale is determined by the residence time of the contaminants in the reactor. It is usual that at least three residence times are required to achieve effective stabilization. Therefore if sampling has been performed from a CSTR, then sampling should have only begun after at least three hydraulic residence times had occurred, and testing continued for at least an additional three residence times to ensure that the aggregate data set is representative of process performance)	☐	☐
3.9	Are samples representative of known, measured and appropriate operating conditions? (Note: this includes technologies that operate on short cycles and so have start and stop cycles which affects the operation of the technology). If the operating conditions are not vital but are recommended, then the reviewer must evaluate operating conditions,	☐	☐
3.10	Were samples and data prepared or provided by a third party? (Note: In some cases, where the expertise rests with the applicant, an independent unbiased third party should witness and audit the collection of information and data about the technology. The witness auditor must not have any vested interest in the technology.)	☐	☐
3.11a-c	**Verification Study Design is Acceptable** **Namely:**		
3.11a	**The samples have been collected when the technology was operated under controlled and monitored conditions, and not at random.**	☐	☐
3.11b	**A verification study design should have been established prior to the test to ensure that the data were collected using a systematic and rational approach**	☐	☐
3.11c	**Verification Study Design should have defined the acceptable values or ranges of values for key operating conditions, and the data collection and analysis methodology**	☐	☐

Once we have found the study design to be acceptable, we move on to our second data checklist. But what if the study design does not meet all of the items mentioned in Table 6? Then we must use our best professional judgment in determining whether the design of the data collection methodology used may result in biased data.

We must also have a good understanding about the samples themselves. Where were they taken, were they grab or composite samples. Were they flow- or time- proportionate. Again, to assist us, we go through a checklist as summarized in Table 7. The answers to the questions will give us a sense of how good (representative) the samples are.

Table 7: Data Validity Checklist

Ref.	Criteria	Meets Criteria	
		Yes	**No**[10]
4.1	Were appropriate sample collection methods used (e.g. random, judgmental, systematic etc?). For example: simple grab samples are appropriate if the process characteristics at a sampling location remain constant over time. Composites of aliquots instead may be suitable for flows with fluctuating process characteristics at a sampling location. Note: Sampling methods appropriate for specific processes may sometimes be described in federal, provincial or local monitoring regulations	☐	☐
4.2	Were apparatus and/or facilities for the test(s) adequate for generation of relevant data? (i.e. testing was performed at a location and under operating conditions and environmental conditions for which the performance claim has been defined.)	☐	☐
4.3	Were operating conditions during the test monitored and documented and provided?	☐	☐
4.4	Has the information and or data on operating conditions and measuring equipment measurements and calibrations been supplied to the Verification Entity?	☐	☐
4.5	Were acceptable protocols used for sample collection, preservation and transport (acceptable protocols include those developed by a recognized authority in environmental testing such as a provincial regulatory body, ASTM, USEPA, Standard Methods)?	☐	☐

[10] Provide written explanations for yes or no meets criteria.

4.6	Were Quality Assurance/Quality Control (QA/QC) (e.g. use of field blanks, standards, replicates, spikes etc) procedures followed during sample collection? A formal QA/QC program, although highly desirable, is not essential, if it has been demonstrated by the vendor's information that quality assurance has been applied to the data generation and collection.	☐	☐
4.7	Were samples analyzed using approved analytical protocols? (e.g. samples analyzed using a protocol recognized by an authority in environmental testing such as Standard Methods, EPA. ASTM etc. Were the chemical analyses at the site in conformance with the SOPs (Standard Operating Procedures) ?	☐	☐
4.8	Were samples analysed within recommended analysis times (especially for time sensitive analysis such as bacteria)	☐	☐
4.9 a-e	Were QA/QC procedures followed during sample analysis Including?	☐	☐
4.9a	Maintaining control charts	☐	☐
4.9b	Establishing minimum detection limits,	☐	☐
4.9c	Establishing recovery values	☐	☐
4.9d	Determining precision for analytical results	☐	☐
4.9e	Determining accuracy for analytical results	☐	☐
4.10 a-c	Was a chain-of-custody (full tracing of the sample from collection to analysis) methodology used for sample handling and analysis. Namely:	☐	☐
4.10a	Are completed and signed chain-of-custody forms used for **each sample** submitted from the field to the analytical lab provided for inspection to the Verification Entity?	☐	☐
4.10b	Are completed and easily readable field logbooks available for the VE to inspect?	☐	☐
4.10c	Are their other chain-of-custody methodology actions and documentation recorded/available (e.g. sample labels, sample seals, sample submission sheet, sample receipt log and assignment for analysis)	☐	☐
4.11	**Experimental Data Set is Acceptable** **(the quality of the data submitted is established using the best professional judgment of the VE)**	☐	☐

5.1.1 Comments for Table 7

We note that an independent accredited laboratory provided the analytical results. We are also satisfied that the applicant presented sufficient quantity and quality of data (except Table 3, data set # 3), which were developed according to reasonable procedures.

Only when we are comfortable with the data quality are we ready to proceed with the statistical evaluation of the claim(s). Let's identify the data set that we will use in our evaluation.

5.2 Data Analysis

Table 1 summarizes the raw data sets that we will use.

Table 1: Performance Data Set # 1

Sample Date	BOD_5 (mg/L)		TSS (mg/L)		Sample Type
	Influent	Effluent	Influent	Effluent	
17-Jan-96	153	16	196	19	Grab
16-Feb-96	242	15	162	20	Grab
17-Mar-96	247	17	179	17	Grab
17-Apr-96	243	18	183	24	Grab
17-May-96	231	15	160	14	Grab
17-Jun-96	203	18	142	7	Grab
17-Jul-96	163	18	170	16	Grab
16-Aug-96	136	17	184	10	Grab
16-Sep-96	147	17	213	18	Grab
16-Oct-96	261	20	185	9	Grab
16-Nov-96	135	18	174	17	Grab
16-Nov-96	227	18	154	17	Grab

As identified in the course of our data quality evaluation, the data were judged to be of adequate quality (see Table 7) to proceed to the next verification step, the statistical assessment of the data. Here we establish within a statistically defined certainty whether or not the technology claim(s) made are supported by the data.

6.0 STATISTICAL EVALUATION OF CLAIM

6.1 Introduction

We now conduct a series of statistical tests to determine the validity of the performance claim. There are 2 questions that we need to answer for both the BOD_5 and TSS effluent data:

1. are the data, effluent BOD_5 as well as effluent TSS ***normally distributed***, and
2. are the ***mean***, effluent BOD_5 as well as effluent TSS, equal to at least 20 and 25 mg/L, respectively, under the specified conditions.

To answer the first question we use Statistical Analysis Worksheet #1 (SAW #1) twice, once for the BOD_5 effluent data and then we go through the same procedure with the effluent TSS data.

To answer the second question we select SAW # 5. Using SAW #5 allows us to test statistically if the mean effluent BOD_5 is equal to 20 mg/L. Again we use it twice, once for the BOD_5 effluent data and then we go through the same procedure again with the effluent TSS data.

So let's look at the data in terms of its distribution (SAW # 1). First we take the data and using Excel® Worksheet, calculate the mean, median and standard deviation for the BOD_5 and TSS data. Actually, for the purposes of the claim we are only interested in the effluent data. In doing the same calculation for the influent, we get a better feel for the overall variability of all data.

Table 8: Performance Data Set # 1

Sample Date	BOD_5 (mg/L)		TSS (mg/L)		Sample Type
	Influent	**Effluent**	*Influent*	**Effluent**	
17-Jan-96	153	16	196	19	Grab
16-Feb-96	242	15	162	20	Grab
17-Mar-96	247	17	179	17	Grab
17-Apr-96	243	18	183	24	Grab
17-May-96	231	15	160	14	Grab
17-Jun-96	203	18	142	7	Grab
17-Jul-96	163	18	170	16	Grab
16-Aug-96	136	17	184	10	Grab
16-Sep-96	147	17	213	18	Grab
16-Oct-96	261	20	185	9	Grab
16-Nov-96	135	18	174	17	Grab
16-Dec-96	227	18	154	17	Grab
n =	**12**	**12**	**12**	**12**	
Mean =	**199.0**	**17.25**	**175.17**	**15.67**	
Median =	**215.0**	**17.50**	**176.50**	**17.00**	
Std. dev. =	**48.56**	**1.42**	**19.30**	**4.91**	

Judging from the results of our spreadsheet analysis of data set #1, we conclude that the data can be used to verify the claim.

6.2 Statistical Analysis Worksheet No. 1 Assessing Normality of Data

This procedure is used to determine if the data variable is normally distributed or log-normally distributed. This is important as the assumption of normality is often invoked in subsequent calculations.

Assumptions:

The xi observations constituting the data set are independent[11].

Data Description	
Parameter: **Effluent BOD_5**	Units: mg/l
Data Location:	**Table 8**
Filename and Location	o electronic database

Determining Potential Normality of Distribution	
Data points may be any real number and the range of possible values is infinite. This is often not the case for a measured value such as a concentration, which cannot be negative. In this case it is sufficient that the majority (95%) of the points lie within 3 standard deviations of the mean of the measured points.	☑ True
The data points are not proportions, rates or frequencies.	☑ True
The data points are not counts.	☑ True
Is the mean approximately the same as the median? median = **17.50** mean = **17.25**	☑ True
Based on guidelines above, the sample is potentially normally distributed.	☑ True ☐ False
If the sampling distribution is potentially normal, and there are more than 10 data points, prepare a normal probability plot of the raw data	

Preparation of Normal Probability Plot	
Order the data (x_i) from smallest to largest. Subsequent calculations use the ordered data.	
Sample size:	n: **12**
Calculate "Blom" coefficients. $p_i = \frac{i - 3/8}{n + 1/4}$, for i = 1 … n.	p_i: see spreadsheet, Table 9
Convert "Blom" coefficients to y_i. $y_i = \sqrt{-\ln(4p_i(1-p_i))}$, for i = 1 … n.	y_i: unnecessary to present the n coefficients here. Attach a table or spreadsheet.
Calculate normal scores. $z_i = sign(p_i - 1/2) \bullet 1.238 \bullet y_i \bullet (1 + 0.0262 y_i)$, for i = 1 … n, where *sign* $(p_i\text{-}1/2)$= -1, for $(p_i\text{-}1/2)$<0, *sign*$(p_i\text{-}1/2)$= +1 for $(p_i\text{-}1/2)$>0, and *sign*$(p_i\text{-}1/2)$=0 for $(p_i\text{-}1/2)$=0.	z_i: unnecessary to present the n coefficients here.
Plot the ordered data against the normal score data.	

[11] A non-rigorous definition of independence is in Appendix A.

Table 9: Spreadsheet for Normality Calculations - BOD_5 effluent data

(similar spreadsheet for Effluent TSS data)

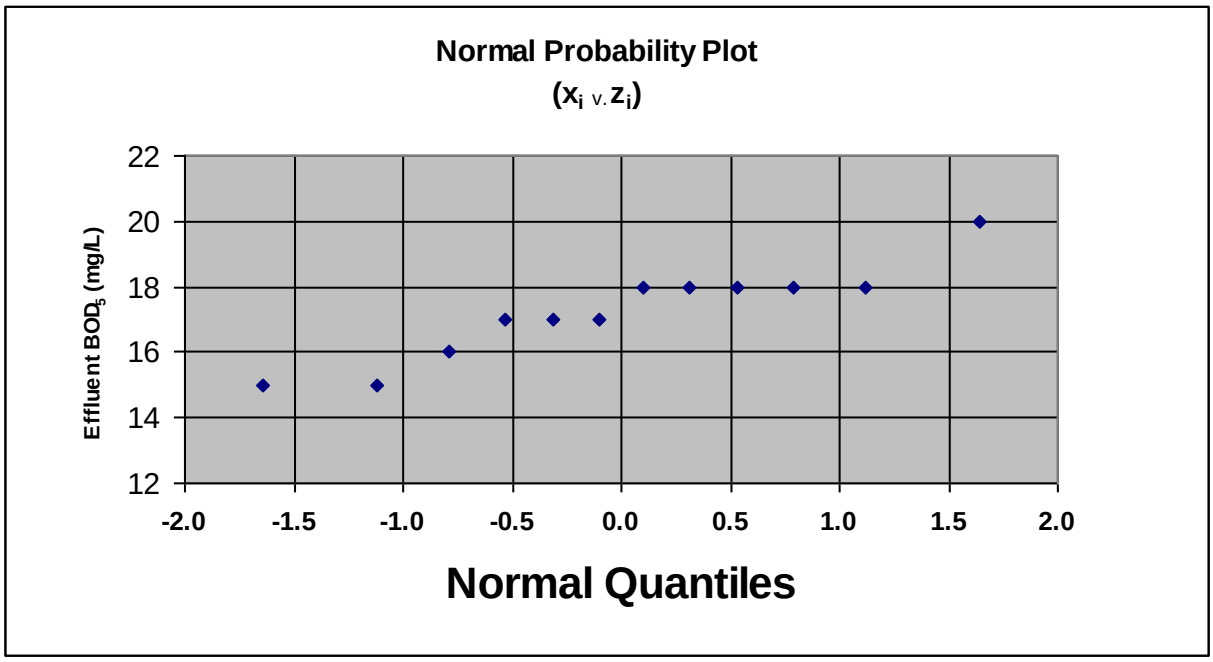
BOD_5 Effluent	Sorted		Index	Coefficients		Normal Scores		Cross-Products
	-	Squared		Blom	Converted Blom	-	Squared	
x_i	$\mathbf{x_i}$	x_i^2	i	p_i	y_i	$\mathbf{z_i}$	z_i^2	$x_i * z_i$
16	**15**	225	1	0.051	1.281	**-1.639**	2.688	-24.592
15	**15**	225	2	0.133	0.881	**-1.116**	1.245	-16.736
17	**16**	256	3	0.214	0.629	**-0.791**	0.626	-12.659
18	**17**	289	4	0.296	0.427	**-0.534**	0.286	-9.085
15	**17**	289	5	0.378	0.249	**-0.310**	0.096	-5.268
18	**17**	289	6	0.459	0.082	**-0.101**	0.010	-1.725
18	**18**	324	7	0.541	0.082	**0.101**	0.010	1.826
17	**18**	324	8	0.622	0.249	**0.310**	0.096	5.578
17	**18**	324	9	0.704	0.427	**0.534**	0.286	9.619
20	**18**	324	10	0.786	0.629	**0.791**	0.626	14.242
18	**18**	324	11	0.867	0.881	**1.116**	1.245	20.084
18	**20**	400	12	0.949	1.281	**1.639**	2.688	32.789
n	**12**							
average	**17.25**							
median	**17.50**							
sum	**207**					**0.000**		**14.073**
sum of squares		**3593**					**9.901**	

SS_x= 22.250
SS_z = 9.901
SS_{xz} = 14.073
W = 0.899

u = 2.485
v = 0.910
u.hat = -2.929
sigma.hat = 0.572
Z.prime = 1.114

We now plot x_i versus z_i (Figure 2). This is our frequency distribution of the effluent BOD_5 data.

Figure 2: Frequency Distribution of BOD_5 Effluent Data

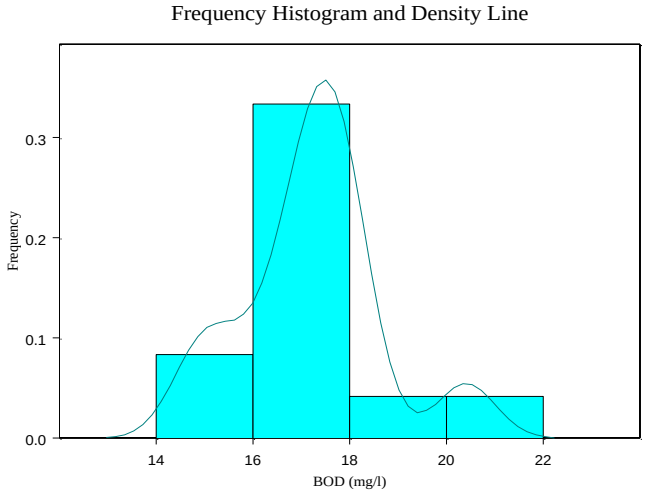


Supporting Documentation, SAW #1

The data do not appear to be normally distributed due to the lack of fit in the tails of the distribution. However, a formal test of normality (the Shapiro-Francia test of composite normality) shows that the BOD_5 effluent data set is normally distributed.

As an aid in understanding the results of the test of normality, the associated frequency histogram and kernel density is shown next. These plots are not part of the minimum requirements for completing a test of normality.

Figure 3: BOD_5 Frequency Histogram



This visualization of the distribution of BOD_5 measurements shows why the formal test of normality did not reject the null hypothesis of composite normality. The data set is reasonably "bell-shaped" and symmetric, given the sample size of 12.

Q_1. Does the data appear to fall on a straight line? ☑ Yes o No

If yes, proceed to test of normality.
If "tails" of distribution fall off the straight-line, log-transform the data and re-plot.

Q_2. Does the log-transformed data appear to fall on a straight line? o Yes o No

If yes, proceed to test of normality. If no, use a test that does not assume normality.

The following are the functions used, as demonstrated in the spreadsheet (Table 9), and now summarized as shown.

Test of Normality	
Estimate the Test Statistic	
$SS_{xz} = \sum_{i=1}^{n} x_i z_i - \left[\left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} z_i \right) / n \right]$	SS_{xz}: **14.073**
$SS_x = \sum_{i=1}^{n} x_i^2 - \left[\left(\sum_{i=1}^{n} x_i \right)^2 / n \right]$	SS_x: **22.250**
$SS_z = \sum_{i=1}^{n} z_i^2 - \left[\left(\sum_{i=1}^{n} z_i \right)^2 / n \right]$	SS_z: **9.901**
Estimate Shapiro-Francia W. $W = \frac{SS_{xz}^2}{SS_x SS_z}$	W: **0.899**
Apply Box-Cox Transformation	
u = ln(n)	u: **2.485**
v = ln (u)	v: **0.910**
$\hat{\mu} = -1.2725 + 1.0521(v - u)$	$\hat{\mu}$: **-2.929**
$\hat{\sigma} = 1.0308 - 0.26758(v + 2/u)$	$\hat{\sigma}$: **0.572**
Transform W to Z'. $Z' = \frac{\ln(1-W) - \hat{\mu}}{\hat{\sigma}}$	Z': **1.114**
If Z' > 1.645 we reject the null hypothesis that the data is normally distributed at the 95% level of confidence. The data is not normally distributed.	

Q_3. Does the data pass a goodness of fit test for normality? ☑ Yes o No
If answers to Q_1 or Q_2 and Q_3 are yes, the raw (or log-transformed) data are normally distributed.
The raw data are Normally Distributed? ☑ Yes o No
The log-transformed data are Normally Distributed? o Yes o No

6.3 Statistical Analysis Worksheet No. 5 Testing Mean is Equal to a Specified Value

H_o: $\mu_1 = \mu_o$

This test is used to determine at a level of 95% confidence that the mean is not equal to some pre-specified value, μ_o. *The value μ_o will often be the performance that a technology is claiming to achieve.*

Assumptions:

- *Data set is normally distributed.*
- *The x_i observations constituting the data set are independent.*

Data Description and Tests of Assumptions	
Parameter: **Effluent BOD_5**	Units: mg/l
Data Location:	**Table 8**
Filename and Location	o electronic database
Based on SAW #1, the data set is normally distributed.	☑ Yes
Common Calculations	
Estimate of μ_1	$\bar{x}$: **17.25**
Hypothesized value μ_o	μ_o: **20**
Sample size n	n: **12**
Estimate of σ^2 $s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right)$	s^2: **2.0227**
If n < 30, the test statistic t, is given by: $t = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}$	t : **-6.6982**
If n ≥ 30, the test statistic Z, is given by: $Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	Z :
Calculations Case A - H_a: $\mu_1 \neq \mu_o$	
If n ≥ 30, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP	critical value: **1.960**
If n < 30, obtain $t_{0.975, n-1}$ from Table B2, Appendix B, GVP	critical value:
Calculations Case B - H_a :$\mu_1 < \mu_o$	
If n ≥30, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: **-1.645**
If n <30, obtain $t_{0.95, n-1}$ from Table B2, Appendix B, GVP, and multiply by -1.	critical value: **-1.796**
Calculations Case C - H_a :$\mu_1 > \mu_o$	
If n ≥30, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: **1.645**
If n <30, obtain $t_{0.95, n-1}$ from Table B2, Appendix B, GVP.	critical value:

Decision Rule
Inferences Case B:

If the test statistics, t or Z ≤ critical value we reject the null hypothesis and accept the alternative hypothesis.

Null Hypothesis:	o **Not Rejected**	☑ **Rejected**
Alternative Hypothesis:	☑ **Accepted**	o **Not Accepted**

Table 10 Spreadsheet for SAW # 5 - Calculation of Variance, Effluent BOD_5

BOD_5 (mg/L)	**Effluent x_i**	**x_i^2**
	16	256
	15	225
	17	289
	18	324
	15	225
	18	324
	18	324
	17	289
	17	289
	20	400
	18	324
	18	324
sum x_i^2		**3593**
sum x_i	**207**	
n	12	
1/(n -1)	0.0909	

variance = $s^2 = 1/(1-n)*[\text{sum } x_i^2 - (\text{sum } x_i)^2/n]$

(see SAW #5)

s^2 = 2.0227

We go through the same process for our effluent TSS data.

The results of the statistical analyses performed are summarized in Table 11. The statistical tests performed support, with 95% confidence, the claims made regarding effluent BOD_5 and effluent TSS produced by the Enviro-Master Biofilter process.

Table 11 Results of Statistical Analyses

Operating Variable EFFLUENT	**Minimum Observed** (mg/L)	**Maximum Observed** (mg/L)	**Mean[1]** (mg/L)	**Median[1]** (mg/L)	**Standard Deviation σ,** (mg/L)	**CLAIM** (mg/L)
BOD_5	15	20	17	18	1.42	20
TSS	7	24	16	17	4.91	25

[1] **values are rounded**

The 95% confidence interval can be determined using the following two equations when n<30. (See SAW #2 for detail)

Lower Confidence limit:

LCL = $\bar{x} - t_{0.975,n-1} \, s/\sqrt{n}$

Upper Confidence limit:

UCL = $\bar{x} + t_{0.975,n-1} \, s/\sqrt{n}$

where "t" = a factor known as student's "t"
s = standard deviation
n = number of samples
x bar = mean

Some values for 't' for various numbers of measurements at the 95% confidence level are shown in Table 12. For 12 observations (n = 12) with the degrees of freedom equals to 11, the 't' value is **2.20**.

Table 12: Student's 't' Values

n	**2**	**3**	**4**	**5**	**6**	**10**	**11**	**12**	**13**	**14**	**15**	**20**
t	12.7	4.3	3.2	2.8	2.6	2.22	2.20	2.18	2.16	2.15	2.13	2.09

as n $\rightarrow$ 100 t $\rightarrow$ 2.0

Table 13 is a summary of the data sets used and the result of the claim verification.

Table 13: Summary of Claim Evaluations

Parameter	**Data Sets Used**	**Claim**	**Claim Verified**
BOD_5 Effluent, (mg/L)	Table 1	**20**	**Yes**
TSS Effluent (mg/L)	Table 1	**25**	**Yes**

7.0 CONCLUSIONS

The performance claim, made by the applicant for their Enviro-Master Biofilter for the treatment of household wastewater effluent from a septic tank, **is validated** in its restated form as follows:

1. **The Enviro-Master Biofilter reduces BOD_5 to at least 20 mg/L under the specified operating conditions (i.e. influent BOD_5 < 200mg/l).**

2. **The Enviro-Master Biofilter reduces TSS to at least 25 mg/L under the specified operating conditions (i.e. influent TSS < 200 mg/L).**

7.1 Limitation of Verification

It is now appropriate to make a statement regarding the limitations of this claim verification. The following is an example of such a statement:

The Canadian ETV Program and the Verification Entity believe that the vendor's technology can achieve the performance levels set out in the Verification Report. This belief is based on the VE's independent analyses of information and declarations provided by the vendor and of samples generated for the performance of the Enviro-Master Biofilter, using verification protocols authorized for the ETV Program. No additional bench or field tests were carried out by the VE to corroborate the data provided. This verification is also based on a use of the technology in accordance with the operating conditions specified by the technology vendor.

The Government of Canada, the Canadian ETV Program and the VE make no express or implied guarantee or warranty as to the performance the **Enviro-Master Biofilter** technology. Nor do they guarantee or warrant this technology to be free from any defects in workmanship, or the integrity or safety of the technology as a whole or it's compliance with such governmental codes, standards and regulations as may be applicable.

7.2 Lessons from Verification

Case Study #1 has been used to demonstrate the following points:

- The SAWs used in this case study describe how to assess the normality of data (SAW #1) and to test if a mean value is equal to a specified value (SAW #5).

- An illustration of how a performance claim must be changed to create a statistically testable hypothesis is given.

- A performance claim may be rejected, revised and then accepted.

It should be emphasized that a hypothesis might be tested in several ways. The method of choice is the simplest, and the most defensible. If test assumptions are met and the data set is "well behaved", contending methods should produce the same conclusions.

A prudent choice of experimental design may reduce costs, increase statistical power and be more defensible than an ill contrived experiment.

Appendix A- ASSUMING DATA IS NOT NORMALLY DISTRIBUTED

Statistical Analysis Worksheet No. 8 Testing Median is Equal to a Specified Value: Effluent BOD_5

H_o: median = m_o

This test is used to determine at a level of 95% confidence that the median is not equal to some pre-specified value, m_o. *The value, m_o will often be the performance that a technology is claiming to achieve. The test presented is the Wilcoxon Signed Ranks test.*

Assumptions:

- *The x_i observations constituting the data set are independent.*
- *The distribution of each d_i is symmetric.*

Data Description and Tests of Assumptions	
Parameter: **Effluent BOD_5**	Units: **mg/L**
Data Location	**Table 8**
Filename and Location	
Based on SAW #1, the data set is **not** normally distributed.	☑ **True**
From a frequency histogram for visual assessment, the data is symmetric	☑ **True**

Common Calculations	
Pre-specified value, m_o.	m_o: **20**
Sample size n	n: **12**
Sort the x_i from smallest to largest.	
Calculate the vector d_i $d_i = m_o - x_i$	**See Attached Spreadsheet Column 3**
Rank the $\|d_i\|$ from smallest to largest to obtain a vector R_i of length n. Identical $\|d_i\|$ are assigned the average of the ranks they would otherwise have received.	**See Attached Spreadsheet Column 5**
The test statistic T^+ is $T^+ = \sum_{i=1}^{n} R_i$, for positive d_i only	T^+: **78**
Calculations Case A - H_a: median ≠ m_o	
Obtain $w_{0.025}$ from Table B4, Appendix B, GVP.	critical value $w_{0.025}$: **14**
critical value = $w_{0.975}$ = n(n + 1)/2 - $w_{0.025}$	critical value $w_{0.975}$: **64**
Calculations Case B - H_a: median < m_o	
Obtain $w_{0.05}$ from Table B4, Appendix B, GVP.	critical value $w_{0.05}$: **18**
critical value = $w_{0.950}$ = n(n + 1)/2 - $w_{0.05}$	critical value $w_{0.95}$: **60**
Calculations Case C - H_a: median > m_o	
Obtain $w_{0.05}$ from Table B4, Appendix B, GVP.	critical value $w_{0.05}$: **18**

Decision Rule

Inference Case B:

If $T^+ \geq w_{0.95}$ we reject the null hypothesis and accept the alternative hypothesis. The median is $< m_o$.

Null Hypothesis:	o **Not Rejected**	☑ **Rejected**
Alternative Hypothesis:	☑ **Accepted**	o **Not Accepted**

CALCULATIONS FOR SAW # 8- EFFLUENT BOD_5

From SAW # 8 m_o = **20** H_o: median = m_o

EFFLUENT BOD_5

①	②	③	④	⑤	⑥	⑦		
x_i	**x_i sorted**	**$d_i = m_o - x_i$**	**abs d_i**	**abs d_i sorted**	**rank**	**T^+**	from table	B4
16	15	5	5	0	1	1	$W_{0.05}$ =	**18**
15	15	5	5	2	2	2	$W_{0.95}$ =	**60**
17	16	4	4	2	3	3		
18	17	3	3	2	4	4		
15	17	3	3	2	5	5		
18	17	3	3	2	6	6		
18	18	2	2	3	7	7		
17	18	2	2	3	8	8		
17	18	2	2	3	9	9		
20	18	2	2	4	10	10		
18	18	2	2	5	11	11		
18	20	0	0	5	12	12		
					Total of T^+	**78**		

Statistical Analysis Worksheet No. 8 Testing Median is Equal to a Specified Value: Effluent TSS

H_o: median = m_o

This test is used to determine at a level of 95% confidence that the median is not equal to some pre-specified value, m_o. *The value, m_o will often be the performance that a technology is claiming to achieve. The test presented is the Wilcoxon Signed Ranks test.*

Assumptions:

- *The x_i observations constituting the data set are independent.*
- *The distribution of each d_i is symmetric.*

Data Description and Tests of Assumptions	
Parameter: **Effluent TSS**	Units: **mg/L**
Data Location	**Table 8**
Filename and Location	
Based on SAW #1, the data set is **not** normally distributed.	☑ **True**
From a frequency histogram for visual assessment, the data is symmetric	☑ **True**

Common Calculations	
Pre-specified value, m_o.	m_o: **25**
Sample size n	n: **12**
Sort the x_i from smallest to largest.	
Calculate the vector d_i $d_i = m_o - x_i$	**See Attached Spreadsheet Column 3**
Rank the $\|d_i\|$ from smallest to largest to obtain a vector R_i of length n. Identical $\|d_i\|$ are assigned the average of the ranks they would otherwise have received.	**See Attached Spreadsheet Column 5**
The test statistic T^+ is $T^+ = \sum_{i=1}^{n} R_i$, for positive d_i only	T^+: **78**
Calculations Case A - H_a: median ≠ m_o	
Obtain $w_{0.025}$ from Table B4, Appendix B, GVP.	critical value $w_{0.025}$: **14**
critical value = $w_{0.975} = n(n+1)/2 - w_{0.025}$	critical value $w_{0.975}$: **64**
Calculations Case B - H_a: median < m_o	
Obtain $w_{0.05}$ from Table B4, Appendix B, GVP.	critical value $w_{0.05}$: **18**
critical value = $w_{0.950} = n(n+1)/2 - w_{0.05}$	critical value $w_{0.95}$: **60**
Calculations Case C - H_a: median > m_o	
Obtain $w_{0.05}$ from Table B4, Appendix B, GVP.	critical value $w_{0.05}$: **18**

Decision Rule

Inference Case B:

If $T^+ \geq w_{0.95}$ we reject the null hypothesis and accept the alternative hypothesis. The median is $< m_o$.

Null Hypothesis:	o **Not Rejected**	☑ **Rejected**
Alternative Hypothesis:	☑ **Accepted**	o **Not Accepted**

CALCULATIONS FOR SAW # 8 - EFFLUENT TSS

From SAW # 8 m_o = **25** H_o: median = m_o

EFFLUENT TSS

①	②	③	④	⑤	⑥	⑦		
x_i	**x_i sorted**	**$d_i = m_o - x_i$**	**abs d_i**	**abs d_i sorted**	**rank**	**T^+**	from table	B4
19	7	18	18	0	1	1	$W_{0.05}$ =	**18**
20	9	16	16	2	2	2	$W_{0.95}$ =	**60**
17	10	15	15	2	3	3		
24	14	11	11	2	4	4		
14	16	9	9	2	5	5		
7	17	8	8	2	6	6		
16	17	8	8	3	7	7		
10	17	8	8	3	8	8		
18	18	7	7	3	9	9		
9	19	6	6	4	10	10		
17	20	5	5	5	11	11		
17	24	1	1	5	12	12		
					Total of T^+	**78**		

Environmental Technology Verification

General Verification Protocol

Appendix F

Case Study 2

February 2007
(Rev. May 2013)

Environment Canada Environnement Canada

Performance Claim Evaluation of the

KN – S8 BIOSLURRY REACTOR

Case Study 2

Prepared for

Canadian ETV Program

By

Pollutech International Limited

Revised by the Canadian ETV Program on:
June 2002
&
February 2007
(Rev. May 2013)

TABLE OF CONTENTS

1.0 Introduction 1
1.1 Literature Background 1
1.2 Technical Specifications of the KN S8 Process 2
2.0 Performance Claim 3
2.1 Operating Characteristics 3
2.2 Verification Experiment 3
2.2.1 Monitoring Data 4
2.2.2 Nutrient Data 4
2.2.3 Sampling Protocols 5
2.2.4 QA/QC 6
2.3 Results 7
3.0 Literature Citations 10
4.0 Statistical Evaluation of the Performance Claim 11
4.1 Claim to Be Verified 11
4.2 Hypothesis Formulation 11
4.3 Testing Efficacy of Total PAH Reduction 12
4.3.1 Preliminary: Combining Data Sets 12
4.3.2 Hypothesis Testing 13
4.4 Lessons 15
5.0 Statistical Analysis Worksheets 16
5.1 Sample Calculations for Combining Total PAH Data Sets 16
5.1.1 Assessing Normality – Treatment 1 16
5.1.2 Testing Equality of Variances 17
5.1.3 Testing Equality for Means 18
5.2 Sample Calculations for Testing Null Hypothesis 1 20
5.2.1 Preliminary Evaluation of Variable Distributions 20
5.2.2 5.2.2 Testing Equality of Variances 23
5.2.2 5.2.2 Testing Equality of Variances 23
5.2.3 Testing Proportional Reduction in Means 24
5.2.4 Testing Equality of Means 26

The case study is not intended to endorse a particular product, nor does it constitute a performance claim. The case study is used to provide guidance when statistically testing performance claims in the context of the ETV program. The rationale behind subjective decisions regarding level of significance, definition of small sample sizes and choice of statistical test methods endorsed, are given in the Introduction to the SAWs, Appendix A of the GVP.

LIST OF TABLES

Table 1: Treatment Definition and Number of Samples 3
Table 2: Characteristics of Soils Used in the Treatments 4
Table 3: Textural Data for Treatments 1 and 2 4
Table 4: Analytical Methods 5
Table 5: Six Week Bioslurry Treatment Results 7

1.0 Introduction

This case study describes the analysis of a performance claim made by the manufacturer of a slurry bioreactor. The relevant SAWs and supporting documentation are presented in appendices. Microsoft Excel® was also used for some calculations.

1.1 Literature Background

Slurry bioreactors have the potential to treat a wide range of organic contaminants such as pesticides, fuels, creosote, pentachlorophenol (PCP), polychlorinated biphenyls (PCBs) and polyaromatic hydrocarbons (PAHs). Although more costly than land treatment systems, bioslurry reactors afford the greatest amount of process control and offer the potential for more rapid contaminant breakdown. Volatile emissions can be captured and amendments such as surfactants can be contained. Highly contaminated soil and sludges with contamination levels ranging from 2,500 mg/kg to 250,000 mg/kg have been effectively treated using slurry bioreactors according to vendors of the technology (USEPA, 1990).

Other than work carried out by ReTeC and ECOVA, there are little data published in North America on the performance of slurry bioreactors in remediating coal tar - contaminated soils. As a consequence, the following review focuses primarily upon the work conducted by these two organizations.

ReTeC, in association with the Gas Research Institute, investigated the bioremediation of Manufactured Gas Plant (MGP) waste contaminated soil at the bench-scale using both soil pans and slurry reactors (ReTeC, 1990). The candidate soils selected for investigation were labelled A through G but results were presented for predominantly sandy soils B and F only. Soil B was from a carbureted water gas plant and had a total PAH level of 162 ppm. Soil F was a highly contaminated soil (total PAH of 25,000 ppm) and was from a MGP site that had used coal and/or oil as feed stock. Soil B and Soil F had organic carbon contents of 0.6 % and 16 %, respectively.

A slurry reactor configuration was used to assess the PAH degradation of the two soils. It was operated at a dissolved oxygen concentration of 7 mg/L, a mixing speed of 1500 rpm, a pH between 7 and 7.5, aeration at 1-2 L/min., a nutrient level of 15-25 mg/L and ambient temperature. The total PAH concentration in soil B was reduced from 160 ppm to approximately 6 ppm in over 6 weeks. This represented a removal efficiency of greater than 95 %. Soil F that originally had a total PAH concentration of 25,000 ppm was cleaned to an endpoint of 5000 ppm, a reduction of 80 %. It was hypothesized that the presence of nonaqueous phase liquid (NAPL) in Soil F and the higher organic content of this soil retarded the dissolution of the PAHs into the aqueous phase. This may account for the difference in endpoints observed between the two soils (ReTeC, 1990).

In a subsequent paper (Linz et al., 1990), ReTeC discussed the treatability of two more soils, soils "J" and "D". Soil J was from a MGP site where both coal and oil feed stocks had been used. The soil was composed of 27 % fine materials (silts and clays) by weight and had an organic carbon fraction of 58 % indicating the presence of lampblack. Soil D was from a carbureted water gas site. Soil D had a similar percentage of finer material as Soil J but the organic carbon content of the soil was only 6.5 %. After sixteen weeks of treatment in a slurry bioreactor, the total PAH content of Soil J was reduced from 29,100 ppm to an endpoint of 16,900 ppm. This represented a much higher endpoint than those exhibited by Soils B and F in

previous studies and is thought to be due primarily to soil J's larger organic fraction and consequent affinity for PAHs. Soil D had an initial total PAH concentration of approximately 193 ppm. After 25 weeks, the concentration decreased to approximately 30 mg/kg. Although the soil had an initial concentration close to that of soil B (162 ppm), the endpoint was approximately three times higher. This may be attributable to the higher organic carbon content of Soil D.

The most comprehensive slurry bioreactor study published to date is by ECOVA Corporation, (Jones et al., 1991). The pilot-scale study made use of 64 L EIMCO airlift bioreactors. The soil was balled-milled in its wet state and sieved such that greater than 30 % of the soil was smaller than # 100 mesh. A creosote-contaminated soil (13,000 total PAHs) was treated as a 30 % (w/v) slurry for a period of twelve weeks. The reactors were amended with an enriched inoculum of indigenous PAH degraders and nutrients.
The major findings of their study were that:

- The greatest decline in PAH and total petroleum hydrocarbons (TPH) concentration occurred in the first two weeks of treatment.
- Total % reduction of soil bound PAHs achieved over nine weeks of testing ranged from 70 to 97 %.
- Amendment with surfactant and additional inoculum did not increase percent removal of the PAHs. The surfactant, Tween 80, was used but the concentration was not specified.
- Concentrations of the majority of liquid-phase PAHs in the post treatment samples were below the established method detection limits.
- Air emission monitoring showed that the majority of emissions occurred during the first five days of reactor operation.
- Comminution of large soil particles was believed to contribute to an apparent increase in the levels of PAHs and TPH due to higher hydrocarbon extraction efficiency as a result of increased soil surface area.

1.2 Technical Specifications of the KN S8 Process

The KN S-8 process consists of a Bioslurry Batch Continuously Stirred Tank Reactor (CSTR) where the soil and the reactor contents are mixed with a marine type propeller three blade mixer allowing for complete suspension of all of the soil particles. The reactor accommodates 2 kg - 300 kg dry weight of contaminated soil. The rpm of the reactor is maintained at a consistent speed using a set point optical tachometer. Mixing alone effectively aerates the slurry and maintains dissolved oxygen levels greater than 2 ppm. The reactor configuration allows for approximately 18% w/v slurry. There is a low energy input necessary to maintain the suspension and aeration. Nitrogen and phosphorus are monitored on line and automatically added to the aqueous phase to maintain a nutrient ratio total organic carbon: total kjedahl nitrogen: total phosphorus (TOC:TKN:TP) of 100:10:1. The vessel configuration allows for proprietary inoculum to be used, chemical treatments to be applied, pH adjustments, etc. The reactors are monitored continuously for D.O., pH, and temperature.

2.0 Performance Claim

The KN S-8 Bioslurry batch CSTR treatment process achieves at least 70% destruction of Total PAH's, reduces 2 ring PAH compounds to 5 ppm or less, and reduces 6 ring PAH compounds to less than 75 ppm for sandy loam soils containing up to 2,000 ppm of adsorbed Total PAH's and residual coal tar (presence of coal tar globules) and 5% TOC.

2.1 Operating Characteristics

These destruction levels are achievable provided the system is operated in accordance with the following conditions:

- The contaminated soil is sieved through a No. 10 standard US sieve and homogenized prior to charging the vessel.
- The reactor is charged with an 18 % (w/v) slurry, treating 270 kg dry weight of contaminated soil.
- The reactor is agitated with a propeller type mixer maintaining slurry suspension and D.O. of 2 mg/L or greater.
- Nitrogen and phosphorus are supplemented to ensure a nutrient ratio TOC:TKN:TP of 100:10:1 in the aqueous phase.
- Temperature is maintained at 21°C, and pH between 7 and 8.
- An inoculum derived from steel mill return activated sludge (RAS) is added to the reactor vessel to expedite degradation of the PAH contaminants.

2.2 Verification Experiment

Coal tar - contaminated soil from a coal gasification site near Rockwood, Ontario was used. On site, the excavated soil had been segregated into two piles - one having a "low" concentration of PAHs (230 ppm total PAHs) and the other having residual coal tar - contamination (2,000 ppm total PAHs - "high" soil). Both soils had high gravel content (over 40 % by wt.) and were sieved prior to being treated. The soil with 230 ppm total PAH contamination was sieved through a No. 10 sieve prior to being homogenized as was the coarser soil with residual coal tar.

Low concentration and high concentration soil was thoroughly mixed separately for four hours using a cement mixer so that the variability in the starting PAH concentration could be minimized. The homogenized soil was laid out on a tarp in the form of a rectangle 4-6 centimeters deep and divided into rectangular batches. The soil from each batch was stored in covered flat trays in a refrigerated room to minimize settling and volatilization. Prior to treatment, samples were taken for PAH analysis from each batch to be treated.

For each of the "high" and "low" soil piles, surfactant was added to some of the sub samples prior to treatment. This was done to test the hypothesis that surfactants would improve the degradation of PAHs at high and low initial concentrations of PAH contamination. The sample sizes for the treatments are shown below.

Table 1: Treatment Definition and Number of Samples

Treatment #	Treatment Definition/Soil Matrix	Initial Samples	Final Samples
1	Low PAH (230 ppm total PAH),	7	3

	no surfactant		
2	Low PAH (230 ppm total PAH), surfactant	3	3
3	High PAH (2,000 ppm total PAH), no surfactant	4	3
4	High PAH (2,000 ppm total PAH), surfactant	4	3

Table 2: Characteristics of Soils Used in the Treatments

Parameter	Soil with Adsorbed PAHs	Soil with Residual Coal Tar
Total PAHs, ppm	230	2,000
TOC, %	3.3	5.1
Texture	Sandy Loam	Sandy Loam
TKN, ppm (µg/g)	393	783
TP, ppm (µg/g)	401	453

Soil is expected to become finer during treatment due to the shearing action of the impeller. Textural analyses were performed on the dry soil before and after treatment.

Table 3: Textural Data for Treatments 1 and 2

	Dry Soil Before Treatment				Dry Soil After Treatment			
Treatment	Texture	% Sand	% Silt	% Clay	Texture	% Sand	% Silt	% Clay
1	Sandy Loam	72.9 (1.2)	19.8 (1.2)	7.3 (0.19)	Fine Sandy Loam	54.5	25.5	20
2	Sandy Loam	73.0 (1.2)	19.8 (1.2)	7.2 (0.34)	Sandy Loam	68.5	19.9	11.6

Standard deviations are bracketed.

2.2.1 Monitoring Data

Daily monitoring data include pH, ambient temperature, and D.O. Nutrient levels are continuously monitored in the reactor vessel. D.O. is maintained in excess of 2 ppm through mixing so that the process is not oxygen-limited. The pH is kept between pH 7 and 8.5, which is conducive to mesophyllic growth. Mean temperatures are kept at approximately 21°C.

2.2.2 Nutrient Data

Since the contaminated soil was nitrogen deficient, ammonium chloride (NH_4Cl) was added periodically to the reactors to maintain a TOC:TKN:TP ratio of 100:10:1 in the aqueous phase. The KN S-8 is equipped with automatic nutrient monitoring and chemical feed equipment.

2.2.3 Sampling Protocols

Samples were taken of the initial and final soil to measure PAH removal. These samples were taken using a coning technique to ensure that the samples had a representative particle size distribution.

To ensure representative sampling, the contents of the reactor were allowed to settle and the aqueous phase was decanted off. The remaining soil was centrifuged, air-dried in a fume hood, ground, homogenized, coned and quartered prior to being sub sampled. There should have been no appreciable loss of semi-volatile compounds when drying the 230 ppm total PAH contaminated soil as previous studies indicated that volatilization was insignificant.

Since the soil contaminated with residual coal tar (the 'High" samples) had a high naphthalene content, the soil could not be dried prior to sub sampling since low molecular weight PAHs would be lost due to volatilization. Therefore, to obtain representative sub samples without drying, the centrifuged soil was frozen prior to grinding in a refrigerated room. This allowed the sample to be well mixed and for a more representative sub sample to be taken without appreciable losses of semi-volatile compounds.

Table 4: Analytical Methods

Analysis	Phase	General Description of Method	Reference	Frequency
TOC	Liquid	Dohrmann, UV method	Method 5310 C Clesceri, L.S. et al., 1989	continuously
PAHs	Soil	GC/MS	KN Method	beginning and end
	Liquid	GC/MS	KN Method	periodically and end
particle size distribution	Soil	sieving	McKeague, J.A. (1979)	beginning and end
oil and grease	Soil	Soxhlet, gravimetric	Method 5520 Clesceri, L.S. et al., 1989	continuously
TKN	Liquid	colorimetric	782-86T Technicon	continuously
TP	Liquid	colorimetric	787-86T Technicon	continuously
NH_4- N	Liquid	Berthelot Reaction colorimetric	780-86T Technicon	continuously

Analysis	Phase	General Description of Method	Reference	Frequency
NO_3	Liquid	colorimetric	782-86T Technicon	continuously

2.2.4 QA/QC

Surrogate Recoveries

Six deuterated compounds (naphthalene-d8, acenaphthene-d10, fluorene-d10, phenanthrene-d10, pyrene-d10, and chrysene-d12) were added to each sample to determine the extraction efficiency. These were not true surrogates as they were added to the soil just prior to Soxhlet extraction and, therefore, did not have appreciable time to adsorb to the matrix, as did the non-deuterated PAHs.

Split Samples and Replicates

From each PAH sample submission lot, the accredited laboratory randomly chose one sample to split and run a duplicate analysis on. The PAH concentrations for the majority of split samples were very similar. The two sets of results reported for the split sample were always averaged together and then used in mass balance calculations. The error associated with analysis and sub sampling was generally less than the error between samples.

Replicate samples were taken at every sampling interval. Data following treatment were more consistent than data before treatment possibly due to thorough homogenization and comminution of the soil during treatment.

A blank was tested for each set of soil samples submitted and analysed during each treatment and replicate. Non-detect or trace amounts of PAHs were found in all blank samples. These data confirmed that there was no significant contamination of any solvents with the target compounds.

2.3 Results

Table 5: Six Week Bioslurry Treatment Results

PAH Group	Initial Concentration (ppm)	Final Concentration (ppm)
Treatment #1 230 ppm total PAHs		
2 ring	1.2	0.41
	0.7	0.48
	0.9	0.45
	1.2	
	1.6	
	1	
	2.1	
3 ring	24.2	4.23
	35.5	4.76
	40.8	6.71
	37.1	
	23.1	
	85.8	
	37.1	
4 ring	111.4	8.14
	91.3	10.16
	107	10.72
	120.1	
	79.4	
	140	
	119.9	
5 ring	52.2	22.86
	52.7	28.53
	49.3	30.15
	61.3	
	43.6	
	66.4	
	68.4	
6 ring	26.3	14.84
	26.7	17.19
	25	18.43
	31	
	22.6	
	35	
	36.9	
total	263	50.84
	214	61.12
	207	66.46
	223	
	251	
	169	

PAH Group	Initial Concentration (ppm)	Final Concentration (ppm)
	329	
Treatment #3, 2,000 ppm total PAHs		
2 ring	170	10.3
	127	8.85
	135	8.51
	268	
3 ring	629	113
	554	104
	551	89
	594	
4 ring	810	169
	604	140
	687	123
	733	
5 ring	209	129
	139	114
	177	97
	157	
6 ring	123	78
	76	74
	99	67
	88	
total	1940	499
	1500	441
	1649	383
	1839	
Treatment #2 230 ppm total PAHs		
2 ring	1.3	0.3
	1.6	0.4
	0.9	0.2
3 ring	42.3	5.3
	42.4	8.2
	41.8	4.4
4 ring	137	10
	118	12.6
	122	7.9
5 ring	71.2	27.9
	68.3	38
	62.2	23
6 ring	38.2	15.8
	40.1	20.2
	36	14.2
total	290	59.3
	270	79.3
	263	49.7
Treatment #4, 2,000 ppm total PAHs		

PAH Group	Initial Concentration (ppm)	Final Concentration (ppm)
2 ring	21	7.73
	452	7.33
	169	8.25
	147	
3 ring	490	69.1
	952	66.4
	612	71.6
	610	
4 ring	666	95.5
	831	87.7
	811	97.8
	721	
5 ring	197	114
	214	107
	174	124
	151	
6 ring	118	60.3
	125	59.8
	139	97.4
	125	
total	1691	347
	2547	328
	1903	399
	1754	

3.0 Literature Citations

Bewley, R., Ellis, B., Theile, P., Viney, I. and J. Rees (1989). "Microbial Clean-Up of Contaminated Soil", Chem. Ind. (London), **23**, 778-783, December.

Canadian Council of Ministers of the Environment (CCME) (1989). "Interim Guidelines for PAH Contamination at Abandoned Coal Tar Sites", prep for Waste Management Committee, Toxic Substances Advisory committee CCME, Nov.

Clesceri, L.S. *et al.*, (1989). *Standard Methods For the Examination of Water and Wastewater*, 17th edition, published by the American Public Health Association, American Water Works Association and Water Environment Fed.

Ellis, W.E., J.R. Payne, and G.D. McNabb (1985). Treatment of Contaminated Soils with Aqueous Surfactants, U.S. EPA Report No. EPA/600/2-85/129, PB 86-122561.

Jones, A.B., Brinkmann, M.R. and W.R. Mahaffey (1991). ECOVA Corp., "OER Section 5 : Design and Operating Data Collection, U.S. EPA Best Demonstrated Available Technology (BDAT) Evaluation, Pilot-Scale Testing For Slurry-Phase Bioremediation of PAHs", Sept 25, P.N. 2003.

Linz, D.G., E.F. Neuhauser and A.C. Middleton (1990). "Perspectives on Bioremediation in the Gas Industry", presented at the Environmental Biotechnology Symposium, Knoxville, Tennessee, Oct. 17.

Mahaffey, W.R., G. Compeau, M. Nelson, and J. Kinsella, ECOVA Corporation (1990). "Development of Strategies for Bioremediation of PAH's and TCE", Proceedings of In Situ Bioremediation of Groundwater and Contaminated Soils, WPCF Annual Conference, Washington D.C., Oct. 7-11.

McKeague, J.A. (1979). Manual of Soil Sampling and Methods of Analysis, 2nd ed, Cdn Soil Survey Cmte.

Remediation Technologies, Inc. (Retec) (1990). Biological Treatment of Soils Containing Manufactured Gas Plant Residues, Topical Report (June 1988 - Feb. 1990), prepared for the Gas Research Institute, GRI-90/0117.

USEPA (1990). Engineering Bulletin, "Slurry Biodegradation", prepared by USEPA, EPA/540/2-90/016, September.

4.0 Statistical Evaluation of the Performance Claim

4.1 Claim to Be Verified

The performance claim to be verified should be stated as hypotheses that may be objectively evaluated, using an appropriate statistical methodology. The following overall performance claim was made.

> **"For soils containing up to 2000 ppm of adsorbed PAH's and residual coal tar, the bioslurry treatment system will remove greater than 70 % total PAH's, reduce 2 ring structures below 5 ppm and reduce 6 ring structures below 75 ppm."**

4.2 Hypothesis Formulation

This performance claim makes specific performance claims, with some restrictions on applicability and the hypotheses being tested are stated explicitly. In order to statistically test these hypotheses we reformulate the performance claims as null and alternative hypotheses.

Null Hypothesis 1: The bioslurry treatment system does not remove more than 70 % of total PAHs.

Alternative Hypothesis 1: The bioslurry treatment system removes more than 70 % of total PAHs.

Null Hypothesis 2: The bioslurry treatment system does not reduce 2 ring structures below 5 ppm.

Alternative Hypothesis 2: The bioslurry treatment system reduces 2 ring structures below 5 ppm.

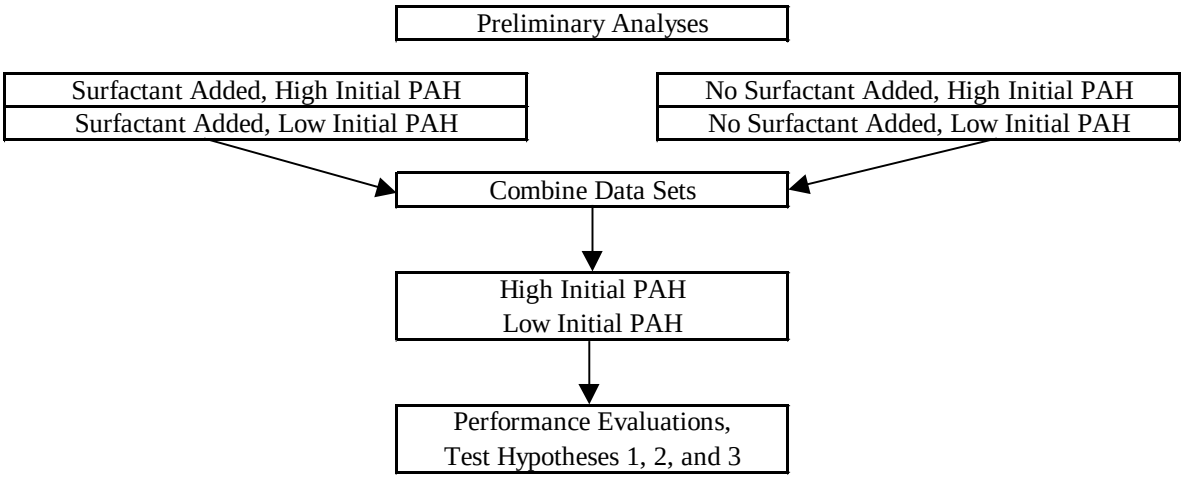
Null Hypothesis 3: The bioslurry treatment system does not reduce 6 ring structures below 75 ppm.

Alternative Hypothesis 3: The bioslurry treatment system reduces 6 ring structures below 75 ppm.

We see that the null hypotheses are stated in a form that may be disproved by sufficient experimental evidence. This follows the paradigm of statistical hypothesis testing: null hypotheses are disproved and alternative hypotheses accepted.

The hypotheses as stated above are one-sided. For example, in the statement regarding total PAHs, we are not interested in the case where the bioslurry may increase the concentration of total PAHs. (From a technical perspective, we know that it is an impossibility that total PAH concentration will increase following biological/mechanical treatment). Thus our hypotheses tests need only consider the case where the treated bioslurry reduces total PAHs.

Statistically, the hypotheses shown above are all of the same form. Therefore only the analysis of the first hypothesis is demonstrated. An outline of the analyses performed is presented below.



4.3 Testing Efficacy of Total PAH Reduction

Prior to testing our hypotheses we note that in the experimental design, a surfactant was added to soils containing low levels (230 ppm) and high levels (2,000 ppm) of total PAHs. The premise was that the addition of a surfactant would improve the degradation of the PAHs during the bioslurry treatment. We begin by testing the hypothesis that the surfactant had no effect. While this is not part of the performance claim, it allows us to pool the data sets that are split or stratified by surfactant type. This increases the statistical power of the tests of hypotheses. (Statistical power is roughly, the power of making a correct decision).

4.3.1 Preliminary: Combining Data Sets

As we are comparing the means between similar soils with and without surfactant added, we use SAW #s 1, 3 and 6. Briefly, SAW # 1 tests the assumption that the two data sets are normally distributed and SAW # 3 tests the assumption that the two variances are equal prior to testing the equality of means using SAW # 6.

We begin by assessing the distribution of total PAHs from each of the 4 treatments, (2 concentrations x presence/absence of surfactant) using SAW #1. As an example, the distribution of data from treatment 1 (low total PAHs, no surfactant added) is assessed in Section 5.1: Sample Calculations for Combining Total PAH Data Sets Assessing Normality, Treatment 1. The distributions from the other 3 treatments are assessed, but are not presented.

Using SAW #1, we see that there is insufficient data to statistically test the hypothesis of normality (See Table 1: Treatment Definition and Number of Samples) but that empirically, we may assume the data is normally distributed. We proceed with the analyses cautiously.

We compare the mean total PAHs between treatments 1 and 2 to test Ho: surfactant has no effect on final total PAH concentration. In the terminology of SAW #6, we are testing H_o: $\mu_1 - \mu_2 = d_o$, where μ_1 = mean of initial total PAHs in treatment 1, μ_2 = mean of final total PAHs in treatment 2 and $d_o = 0$. In the course of using SAW #6, we must test the equality of variances. This is done using SAW #3. Once the test of equality of variances is completed, we resume

work with SAW#6. The results of these analyses are presented in sections 5.1.2: Testing Equality of Variances, Treatment 1, and section 5.1.3: Testing Equality of Means, Treatment 1, and the Case 2 spreadsheet.

The equality of variance test shows that the variances of the two data sets are equal. We may then compare the means using the SAW #6. Analyses using SAW #6 show that the mean total PAH concentration of the two data sets are similar. Therefore we may combine the two data sets.

Note that in Microsoft Excel®, the larger of the two variances must be designated as "Variable 1" when testing the equality of variances. The two variances are equal, as are the mean total PAH concentrations for the soil samples designated as containing "low" amounts of total PAHs. Therefore these two data sets may be combined.

The same analyses were conducted for the treatment of highly contaminated soil without surfactant (treatment 3) and with surfactant (treatment 4). These results are not presented below but may be found in the Case 2 spreadsheet. Again, both means and variances were equal so that these two data sets may be combined.

4.3.2 Hypothesis Testing

We now go on to test null hypothesis 1: The bioslurry treatment system does not remove more than 70 % of total PAHs against the alternative hypothesis 1: The bioslurry treatment system removes more than 70 % of total PAHs. This is a one-sided hypothesis as we are interested in a reduction **greater than** 70% not testing that the reduction in total PAHs is **equal to** 70%. This analysis would ordinarily be conducted twice; once for soils with low levels of total PAHs and again for soils with high levels of PAHs. However as this is a teaching case[1] rather than a performance claim evaluation, only the soils with "low" amounts of total PAHs are tested following this null hypothesis.

The initial and final measurements are not paired. This precludes using the more powerful, paired analogue of the t-test. As a proportional reduction is being claimed rather than an absolute reduction we cannot simply compare the initial and final means as the variance term will be incorrectly estimated[2]. SAW #4 describes a procedure for testing this hypothesis. The results of this analysis are presented in Section 5.2: Sample Calculations for Testing Null Hypothesis 1. The steps in this analysis are:

- Evaluate the distribution of variables. (SAW # 1)
- Check assumption of equality of variances. (SAW # 3)
- Convert initial data by multiplying by the hypothesized proportional reduction in total PAHs. (SAW # 4)

[1] Note that it is possible to conduct a single multivariate test of hypotheses using both "low" and "high" data sets simultaneously. However this procedure may not be more powerful than the univariate tests being endorsed, and consequently will not be used.

[2] Testing H_o: $\mu_2 = p\,\mu_1$ where p represents the proportional reduction in a sample due to the application of a technology. It can be shown that this is equivalent to testing H_o: $u_1(1- p) = \mu_2$. The coefficient 1-p in the hypothesis must be accounted for in tests of hypotheses.

- Check assumption of equality of variances using converted initial data and final data. (SAW # 3)
- Test the equality of converted initial data with final data. (SAW # 6)

We see that the performance claim "The bioslurry treatment system removes more than 70% of total PAHs" cannot be confirmed. The null hypothesis is restated as, Ho: The bioslurry treatment system does not remove more than 60% of total PAHs. Following the methods described above this reformulated hypothesis is tested using the same SAWs. The results of these analyses are not presented but are contained in the accompanying Microsoft Excel® spreadsheet. This analysis produces a t-statistic of 4.572 that is greater than the critical value of 2.145. Therefore we reject the null hypothesis and accept the alternative hypothesis. The performance claim:

The bioslurry treatment system removes more than 60% of total PAHs when the soil loading of total PAHs is "low",

is accepted.

4.4 Lessons

Case study 2 has been used to demonstrate the following items.

- The SAWs used in this case study describe how to assess the normality of data (SAW #1), compare two variances (SAW #3), compare two means with equal variances (SAW #6) and determine the significance of a proportional decrease in two means (SAW #4).
- An illustration of how a performance claim must be changed to create a statistically testable hypothesis is given.
- It was shown that the SAWs do not cover all situations. A subjective, but informed decision regarding the assumption of normality is recommended when sample sizes are < 10.
- It was shown that when possible, data sets should be combined to increase sample sizes.
- A performance claim may be rejected, revised and then accepted.

It should be emphasized that a hypothesis might be tested in several ways. The method of choice is the simplest, and the most defensible. If test assumptions are met and the data set is "well-behaved" contending methods should produce the same conclusions.

A prudent choice of experimental design may reduce costs, increase statistical power and be more defensible than an ill-contrived experiment. For example, in this case study, paired initial and final measurements would reduce the number of chemical analyses made, as the non-paired initial measurements would be unnecessary.

5.0 Statistical Analysis Worksheets

5.1 Sample Calculations for Combining Total PAH Data Sets

5.1.1 Assessing Normality – Treatment 1

SAW # 1 Assessing Normality of Data

This procedure is used to determine if the data variable is normally distributed or log-normally distributed. This is important as the assumption of normality is often invoked in subsequent calculations.

Assumptions:
The y_i observations constituting the data set are independent[3].

Data Description	
Parameter: treatment 1 low total PAHs, no surfactant	Units:ppm
Data Location	o attached page
Filename and Location	☑ electronic database

Rules-of-Thumb for Determining Potential Normality of Distribution	
Data points may be any real number and the range of possible values is infinite. This is often not the case for a measured value such as a concentration, which cannot be negative. In this case it is sufficient that the majority (95%) of the points lie within 3 standard deviations[4] of the mean of the measured points.	☑ True
The data points are not proportions, rates or frequencies.	☑ True
The data points are not counts.	☑ True
Is the mean approximately the same as the median? median = mean =	☑ True
Based on guidelines above, the sample is potentially normally distributed.	☑ True o False
If the sampling distribution is potentially normal, and there are more than 10 data points, prepare a normal probability plot of the raw data	

NOTE: THERE IS INSUFFICIENT DATA TO STATISTICALLY TEST THE HYPOTHESIS OF NORMALITY. THEREFORE, ASSUME DATA IS NORMALLY DISTRIBUTED AND PROCEED CAUTIOUSLY.

[3] A non-rigorous definition of independence is in Appendix A.
[4] Standard deviation is defined in Appendix D.

5.1.2 Testing Equality of Variances

SAW # 3 Testing Equality of Two Variances (Refer to Appendix I – Case Study Excel Spreadsheet for calculation)

H_o: $\sigma_1^2 = \sigma_2^2$

This test is used to determine at a level of 95 % confidence, if two variances are equal. The equality of variances is important when pooling data sets. The formulae presented below are applicable when the two data sets are equal or unequal in number.

Assumptions:

- *Both data sets are normally distributed.*
- *The x_i observations constituting the data set are independent*[5]*.*
- *Data sets are independent of one another*[6]*.*

Data Description and Tests of Assumptions	
Parameter: treatment 1 and 2 total PAH	Units:ppm
Data Location	o attached page
Filename and Location	☑ electronic database
Based on SAW#1,the data sets are normally distributed.	o Yes

Common Calculations	
Estimate of σ_1^2 (Let larger variance correspond to numerator)	s_1^2 : 228.1
Estimate of σ_2^2	s_2^2 : 63.0
Degrees of Freedom Data Set 1 = n_1 - 1	$v_{1:}$ 2
Degrees of Freedom Data Set 2 = n_2 - 1	v_2:2
Test statistic F = σ_1^2 / σ_2^2	F: 3.62
Calculations Case A - H_a: $\sigma^2_1 \neq \sigma^2_2$	
Obtain $F_{\alpha/2, v\,1, v\,2}$ from Table B3, Appendix B, GVP.	critical value:
Calculations Case B - H_a: $\sigma^2_1 > \sigma^2_2$ or H_a: $\sigma^2_1 < \sigma^2_2$	
Obtain $F_{\alpha, v\,1, v\,2}$ from Table B3, Appendix B, GVP	critical value: 19.00

Test stat F (3.62) < critical F (19), we can not reject the null hypothesis and accept the alternative hypothesis, therefore the two variances are equal.

Null Hypothesis $\sigma_1^2 = \sigma_2^2$: ☑ **Not Rejected** **Rejected**

Performance Claim: **Accepted** **Not Accepted**

[5] A non-rigorous definition of independence is in Appendix A

[6] The independence of data sets is defined in Appendix A

5.1.3 Testing Equality for Means

SAW # 6 Testing Equality of Two Means when Sample Variances are Assumed Equal (Refer to Appendix I – Case study Excel Spreadsheet for calculations)

H_o: $\mu_1 - \mu_2 = d_o$

This test is used to determine at a level of 95% confidence, If the difference of two means are equal to a pre-specified difference d_o. If this difference is 0, we are testing that the two means are equal or, $\mu_1 = \mu_2$. The formulae presented below are applicable when the two data sets are equal or unequal in number.

Assumptions:

- *Both data sets are normally distributed.*
- *Variances estimated from both data sets are equal.*
- *The x_i observations constituting the data set are independent*[7]*.*
- *Data sets are independent of one another*[8]*.*

Data Description and Tests of Assumptions		
Parameter: treatment 1 and 2 total PAH	Units: ppm	
Data Location	o attached page	
Filename and Location	☑ electronic database	
Based on SAW#1,the data sets are normally distributed.	o Yes	
Based on SAW #3, the variances are equal.	o Yes	o No use SAW # 7

Common Calculations	
Pre-specified value d_o	d_o:0
Estimate of μ_1	$\bar{x}_1$:59.5
Estimate of μ_2	$\bar{x}_2$:62.8
n_1: 3	n_2:3
Total sample size $n = n_1 + n_2$	n: 6
Estimate of σ_1^2 $s_1^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n_1} x_{1i}\right)^2}{n_1}\right]$	s_1^2 :63.0

[7] A non-rigorous definition of independence is in Appendix A

[8] The independence of data sets is defined in Appendix A

Common Calculations	
Estimate of σ_2^2 $s_2^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n_2} x_{2i}\right)^2}{n_2}\right]$	s_2^2 :228.1
Estimate of pooled variance σ_p^2 $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	s_p^2 :145.54
If $n_1 + n_2 - 2 < 30$, the test statistic t, is given by: $t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t:-0.334
If $n_1 + n_2 - 2 \geq 30$, the test statistic Z, is given by: $Z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	Z:
Calculations Case A - H_a: $\mu_1 \neq \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	Critical value:1.960
If $n < 30$, obtain $t_{0.975, n-2}$ from Table B2, Appendix B, GVP.	critical value: 2.776
Calculations Case B - H_a :$\mu_1 < \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If $n < 30$, obtain $t_{0.95, n-2}$ from Table B2, Appendix B, GVP, and multiply by -1.	critical value:
Calculations Case C - H_a :$\mu_1 > \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645
If $n < 30$, obtain $t_{0.95, n-2}$ from Table B2, Appendix B, GVP.	critical value:

Decision Rule

Inferences Case A:

Test statistic |t| (-0.334) < critical value (2.776) we can not reject the null hypothesis and accept the alternative hypothesis, therefore the two means are equal.

Null Hypothesis: ☑ **Not Rejected** **Rejected**

Alternative Hypothesis: **Accepted** **Not Accepted**

5.2 Sample Calculations for Testing Null Hypothesis 1

5.2.1 Preliminary Evaluation of Variable Distributions

We begin by evaluating the normality of the initial data set for the soils contaminated with "low" concentrations of total PAHs following SAW # 1.

SAW # 1 Assessing Normality of Data (Refer to Appendix I- Case study Excel Spreadsheet for calculation)

This procedure is used to determine if the data variable is normally distributed or log-normally distributed. This is important as the assumption of normality is often invoked in subsequent calculations.

Assumptions:
The x_i observations constituting the data set are independent[9].

Data Description	
Parameter: low concentration total PAH	Units:ppm
Data Location	o attached page
Filename and Location	o electronic database

Rules-of-Thumb for Determining Potential Normality of Distribution	
Data points may be any real number and the range of possible values is infinite. This is often not the case for a measured value such as a concentration, which cannot be negative. In this case it is sufficient that the majority (95%) of the points lie within 3 standard deviations[10] of the mean of the measured points.	☑ True
The data points are not proportions, rates or frequencies.	☑ True
The data points are not counts.	☑ True
Is the mean approximately the same as the median? median = mean =	☑ True
Based on guidelines above, the sample is potentially normally distributed.	☑ True o False
If the sampling distribution is potentially normal, and there are more than 10 data points, prepare a normal probability plot of the raw data	

Preparation of Normal Probability Plot	
Order the data (x_i) from smallest to largest.	
Sample size:	n:10
Calculate "Blom" coefficients. $p_i = \frac{i - 3/8}{n + 1/4}$, for i = 1 … n.	p_i: unnecessary to present the n coefficients here. Attach a table or spreadsheet.

[9] A non-rigorous definition of independence is in Appendix A.
[10] Standard deviation is defined in Appendix D.

Convert "Blom" coefficients to x_i. $x_i = \sqrt{-\ln(4p_i(1-p_i)}$, for i = 1 … n.	x_i: unnecessary to present the n coefficients here. Attach a table or spreadsheet.
Calculate normal scores. $z_i = sign(p_i - 1/2) \bullet 1.238 \bullet x_i \bullet (1 + 0.0262 x_i)$, for i = 1 … n, where sign = -1, for negative values, +1 for positive values and 0, otherwise.	z_i: unnecessary to present the n coefficients here.
Plot the ordered data against the normal score data.	

Q_1. Does the data appear to fall on a straight line? ☑ Yes o No

If yes, proceed to formal test of normality.
If no: If "tails" of distribution fall off the straight-line, log-transform the data and re-plot.

Q_2. Does the log-transformed data appear to fall on a straight line? o Yes o No

If yes, proceed to formal test of normality.
If no, use a test that does not assume normality. For example, SAW # 8 and 9.

Test of Normality	
Estimate the Test Statistic	
$SS_{xz} = \sum_{i=1}^{n} x_i z_i - \left[\left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} z_i \right) / n \right]$	SS_{xz} :384.2783 – 2479 * 0/10 = 384.2783
$SS_x = \sum_{i=1}^{n} x_i^2 - \left[\left(\sum_{i=1}^{n} x_i \right)^2 / n \right]$	SS_x : 633515 – 2479^2 /10 = 18970.9
$SS_z = \sum_{i=1}^{n} z_i^2 - \left[\left(\sum_{i=1}^{n} z_i \right)^2 / n \right]$	SS_z :7.980165949 – 0^2/10 = 7.980165949
Estimate Shapiro-Francia W. $W = \frac{SS_{xz}^2}{SS_x SS_z}$	W: (384.2783/(18970.9*7.980165949)^.5) = 0.975420396
Apply Box-Cox Transformation	
u = ln(n)	u:2.302585093
v = ln (u)	v:0.834032445
$\hat{\mu} = -1.2725 + 1.0521(v - u)$	$\hat{\mu}$:-2.817564241
$\hat{\sigma} = 1.0308 - 0.26758(v + 2/u)$	$\hat{\sigma}$:0.575212563
Transform W to Z′. $Z' = \frac{\ln(1-W) - \hat{\mu}}{\hat{\sigma}}$	Z' :-1.544253557
If Z′ > 1.645 we reject the null hypothesis that the data are normally distributed at the 95% level of confidence. The data are not normally distributed.	

Q_3. Does the data pass a goodness of fit test[11] for normality? ☑ Yes o No

If answers to questions Q_1 or Q_2 and Q_3 are yes, the raw (or log-transformed) data are normally distributed.

The raw data are Normally Distributed?	☑ Yes	o No
The log-transformed data are Normally Distributed?	o Yes	o No

You can now proceed to the next appropriate SAW.

[11] Recommended test of normality for manual calculations is the Royston modification of the Shapiro-Francia test. Users with access to statistical software are advised to use the Shapiro-Wilks test.

5.2.2 Testing Equality of Variances

SAW # 3 Testing Equality of Two Variances

H_o: $\sigma_1^2 = \sigma_2^2$

This test is used to determine at a level of 95 % confidence, if two variances are equal. The equality of variances is important when pooling data sets. The formulae presented below are applicable when the two data sets are equal or unequal in number.

Assumptions:

- *Both data sets are normally distributed.*
- *The x_i observations constituting the data set are independent*[12]*.*
- *Data sets are independent of one another*[13]*.*

Data Description and Tests of Assumptions	
Parameter: low concentration total PAH	Units: ppm
Data Location	o attached page
Filename and Location	o electronic database
Based on SAW#1,the data sets are normally distributed.	☑ Yes

Common Calculations	
Estimate of σ_1^2 (Let larger variance correspond to numerator)	s_1^2 : 189.7
Estimate of σ_2^2	s_2^2 : 119.7
Degrees of Freedom Data Set 1 = n_1 - 1	v_1: 9
Degrees of Freedom Data Set 2 = n_2 - 1	v_2: 5
Test statistic F = σ_1^2 / σ_2^2	F:1.59
Calculations Case A - H_a: $\sigma^2_1 \neq \sigma^2_2$	
Obtain $F_{\alpha/2, v1, v2}$ from Table B3, Appendix B, GVP.	critical value: 4.77
Calculations Case B - H_a: $\sigma^2_1 > \sigma^2_2$ or H_a: $\sigma^2_1 < \sigma^2_2$	
Obtain $F_{\alpha, v1, v2}$ from Table B3, Appendix B, GVP	critical value:

The test stat F (1.59) < the critical value (4.77) we can not reject the null hypothesis and accept the alternative hypothesis, therefore the two variances are equal.

Null Hypothesis $\sigma_1^2 = \sigma_2^2$: **☑ Not Rejected** **Rejected**

Performance Claim: **Accepted** **Rejected**

[12] A non-rigorous definition of independence is in Appendix A

[13] The independence of data sets is defined in Appendix A

5.2.2 Testing Proportional Reduction in Means

Saw # 4 Testing Proportional Reductions
H_o: $\mu_2 = p\mu_1$

This test[14] is used to determine at a level of 95% confidence, whether a pre-specified percentage change occurs in a sample, as the result of applying a process or technology. For example, a claim may state that a technology removes "p%" of contaminant from a process stream (i.e., 95% confident that the technology can remove "p%" of contaminant). If μ_1 is the mean of a sample prior to the application of the technology, we wish to test whether the mean after treatment μ_2 is equal to (1-p%)μ_1. The formulae presented below are applicable when the sizes of both data sets are equal or unequal.

Assumptions:

- *Both data sets are normally distributed.*
- *The x_i observations constituting the data set are independent[15].*
- *Data sets are independent of one another[16].*

Data Description	
Parameter:	Units:
Data Location	o attached page
Filename and Location	o electronic database

Preliminary Calculations and Tests of Assumptions		
Convert the pre-technology observations x_{1i} for i = 1 … n_1 to x_{1i}^* = (1-p%) x_{1i} .		
Based on SAW#1, samples x_{1i}^* and x_{2i} are normally distributed.	☑ Yes	o No
If one or both samples are not normally distributed, use SAW #9 to test the equality of median of the transformed pre-technology observations, x_1^* with the median of the post-technology observations, x_2.		
Based on SAW #3, the variances are equal.	☑ Yes	o No

Common Calculations	
Estimate test statistic, t or Z using SAW #6 if variances are equal or SAW #7 if variances are unequal. Substitute x_{1i}^* for x_{1i} in all calculations.	
Total sample size n = $n_1 + n_2$	n: 16
Calculations Case A - H_a: $\mu_2 \neq (1-p\%)\mu_1$	
If n or dof_e[17] ≥ 30, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	critical value: 1.960
If n or dof_e[17] < 30, obtain $t_{0.975,\ n-2\ or\ dofe}$[18] from Table B2, Appendix B, GVP.	critical value: 2.145
Calculations Case B H_a : $\mu_2 < (1-p\%)\mu_1$	
If n or dof_e[11] ≥ 30, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645
If n or dof_e[11] < 30, obtain $t_{0.95,\ n-2\ or\ dofe}$[18] from Table B2, Appendix B, GVP	critical value:

[14] A more rigorous (but more difficult to implement) test of this hypothesis is provided in Kendall, M. and A. Stuart. 1979. The advanced theory of statistics, Volume 2: Inference and relationship. Chapter 21, pg 152. Charles Griffin and Co. Ltd., London.
[15] A non-rigorous definition of independence is in Appendix A.
[16] The independence of data sets is defined in Appendix A.
[17] For SAW #6, the choice for the use of Z or t is based on n. For SAW #7, the choice for the use of Z or t is based on the effective degrees of freedom, dof_e
[18] For SAW #6, the degrees of freedom are n-2. For SAW #7, the degrees of freedom are dof_e

Calculations Case C $H_a: \mu_2 > (1\text{-}p\%)\mu_1$	
If n or dof_e[11] ≥ 30, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If n or dof_e[11] < 30, obtain $t_{0.95,\ n\text{-}2 \text{ or } dofe}$[12] from Table B2, Appendix B, GVP, and multiply by -1	critical value:

For p=0.3
ItI stat (1.999)[19] < critical value (2.145), we can not reject the null hypothesis and accept the alternative hypothesis, therefore the two means are equal.

Null Hypothesis:	☑ **Not Rejected**	**Rejected**
Performance Claim:	**Accepted**	**Rejected**

[19] From SAW #6 or Appendix I Case Study Excel Spreadsheet.

5.2.3 Testing Equality of Means

SAW # 6 Testing Equality of Two Means when Sample Variances are Assumed Equal

H_o: $\mu_1 - \mu_2 = d_o$

This test is used to determine at a level of 95% confidence, if two means are equal to a pre-specified difference d_o. If this difference is 0, we are testing that the two means are equal or, $\mu_1 = \mu_2$. The formulae presented below are applicable when the two data sets are equal or unequal in number.

Assumptions:

- *Both data sets are normally distributed.*
- *Variances estimated from both data sets are equal.*
- *The x_i observations constituting the data set are independent[20].*
- *Data sets are independent of one another[21].*

Data Description and Tests of Assumptions		
Parameter: low concentration total PAHs	Units: ppm	
Data Location	o attached page	
Filename and Location	o electronic database	
Based on SAW#1,the data sets are normally distributed.	☑ Yes	
Based on SAW #8, the variances are equal.	☑ Yes	o No use SAW # 7

Common Calculations	
Pre-specified value d_o	d_o:0
Estimate of μ_1	$\overline{x}_1$: 74.4
Estimate of μ_2	$\overline{x}_2$:61.1
n_1: 10	n_2:6
Total sample size $n = n_1 + n_2$	n: 16
Estimate of σ_1^2 $s_1^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n_1} x_{1i}\right)^2}{n_1}\right]$	s_1^2 :189.71

[20] A non-rigorous definition of independence is in Appendix D.
[21] The independence of data sets is defined in Appendix D.

Common Calculations	
Estimate of σ_2^2 $s_2^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n_2} x_{2i}\right)^2}{n_2}\right]$	s_2^2:119.7
Estimate of pooled variance σ_p^2 $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	s_p^2:164.7
If $n_1 + n_2 - 2 < 30$, the test statistic t, is given by: $t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t:1.999
If $n_1 + n_2 - 2 \geq 30$, the test statistic Z, is given by: $Z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - d_o}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	Z:
Calculations Case A - H_a: $\mu_1 \neq \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.975}$ from Table B1, Appendix B, GVP.	Critical value:1.960
If $n < 30$, obtain $t_{0.975, n-2}$ from Table B2, Appendix B, GVP.	critical value:2.145
Calculations Case B - H_a :$\mu_1 < \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.05}$ from Table B1, Appendix B, GVP.	critical value: -1.645
If $n < 30$, obtain $t_{0.95, n-2}$ from Table B2, Appendix B, GVP, and multiply by -1.	critical value:
Calculations Case C - H_a :$\mu_1 > \mu_2 + d_o$	
If $n \geq 30$, obtain $Z_{0.95}$ from Table B1, Appendix B, GVP.	critical value: 1.645
If $n < 30$, obtain $t_{0.95, n-2}$ from Table B2, Appendix B, GVP.	critical value:

ItI stat (1.999)[22] < critical value (2.145), we can not reject the null hypothesis and accept the alternative hypothesis, therefore the two variances are equal.

Null Hypothesis: ☑ **Not Rejected** **Rejected**

Performance Claim: **Accepted** ☑ **Rejected**

[22] From SAW #6 or Appendix I Case Study Excel Spreadsheet.

Canadian Environmental Technology Verification (ETV) Program

General Verification Protocol

Appendix G

The Verification Report Format

February 2007
(Rev. May 2013)

Environment Canada Environnement Canada

EXECUTIVE SUMMARY

[VERIFICATION ENTITY INSERT HERE]

TABLE OF CONTENTS

[VERIFICATION ENTITY UPDATE THIS:]

1.0 Introduction

The methodology outlined in the General Verification Protocol enables the reviewer, the VE, to perform a structured and systematic examination of the field test program and its results. A series of checklists is used, so that many items of review are covered efficiently. The checklists, Tables 1 to 8, can be ticked electronically (using the Protect Form on / off key in Word). For the Verification Report, the VE completes the checklists directly in the report. Within each individual question, the VE may add a short explanation for the answer, if needed. More extensive text relating to the topics covered in a particular checklist/Table would be placed adjacent to the Table.

[Within each checklist, there is indication whether the requirement must be fulfilled (mandatory) or is treated as useful or desirable information (optional).]

The following is an example verification report template that may be used to prepare the final verification report. The VE may propose to use other than the following standard format, if prior approval is obtained before drafting the report.

1.1 Report Format

This report contains 6 Chapters. Chapter 1 provides a brief introduction to the ETV Program. Chapter 2 summarizes information provided by the applicant as part of the application pre-screening and the formal application process. Chapter 3 contains the results of an evaluation as to whether the Test Plan, Test Execution and Data from [THE PROPONENT NAME HERE] was adequate to proceed to the next step, the claim evaluation.
In Chapter 4 the claim is evaluated and the results of the statistical analyses performed are presented. Chapter 5 covers the Audit Trail of the verification. This report concludes with Chapter 6, with a statement of the performance claim and verification results supported by the data submitted by [THE PROPONENT NAME HERE] and verified by [VERIFICATION ENTITY NAME HERE] using the Canadian ETV Program General Verification Protocol 2007.

Appendix A contains the worked up detailed Statistical Analyses Worksheets (SAWS) used for the verification. Appendix B contains the SAW tables used. [VERIFICATION ENTITY COMPLETE AS NECESSARY WITH ADDITIONAL APPENDICES TOO]

1.2 Background

The Environmental Technology Verification (ETV) Program is an Environment Canada initiative delivered by Globe Performance Solutions. The ETV Program is designed to support Canada's environment industry by providing credible and independent verification of environmental technology performance claims. Interested suppliers are invited to apply to the program for verification of the claims they make concerning the performance of their environmental technologies. If the claim is verified, the company is entitled to use the ETV Program verification mark along with the accompanying certificate in their marketing activities in Canada and abroad.

In order for a technology to be ***eligible*** for the ETV program, it must meet the following criteria:

1. It must be either:
    a) an environmental technology or process that offers an environmental

benefit or addresses an environmental problem, or

b) an equipment-based environmental service that can make claims based solely on measurable performance of the equipment.

2. The claim must be:
   - specific and unambiguous.
   - meaningful and nontrivial.
   - measurable and verifiable.

3. To be eligible for receipt of a Verification Certificate, the technology must also be currently commercially available or commercially ready for full-scale application.

The Assessment of the [INSERT PROPONENTS TECHNOLOGY NAME HERE] claim made for the [INSERT NAME OF MODEL/VERSION OF THE TECHNOLOGY] system was conducted using the 2007 version of the Canadian ETV Program General Verification Protocol. The evaluation process consisted of three stages: application review, data quality assessment and evaluation of the technology claim made. For a claim to be verified, the following basic criteria had to be fulfilled:

1. The technology is based on sound scientific and engineering principles.
2. The claim is fully supported by peer-review quality data, which are supplied by the applicant.
3. The conditions of performance for the claim are clearly defined.

1.3 Objectives

The objective is to report on the verification of the performance claim made by [INSERT PROPONENT NAME HERE]. for their [INSERT MODEL/TYPE/VERSION NUMBER HERE] system technology. This report summarizes the findings of the Verification Entity, [INSERT VERIFICATION ENTITY NAME HERE], based on the information and data supplied by [INSERT PROPONENT NAME HERE] to Globe Performance Solutions.

1.4 Scope

This verification project was conducted by [INSERT VERIFICATION ENTITY NAME HERE] using the General Verification Protocol, sanctioned by the Canadian ETV Program, dated 2007.

1.5 Legal Notices

1.5.1 Limitations of Verification

The following or similar statements must be included in the Verification Report, accompanied by signatures by authorized personnel in each of the organizations named here.

The Verification Report is written by the Verification Entity (VE), reviewed by Globe Performance

Solutions, modified if required, and approved by Globe Performance Solutions. Globe Performance Solutions issues the Verification Report.

1.5.2 Required Statement by Globe Performance Solutions

Globe Performance Solutions, and its subcontractors, provide the verification services solely on the basis of the information supplied by the applicant, test agency or vendor and assume no liability thereafter. The responsibility for the information supplied remains solely with the applicant or vendor and the liability for the purchase, installation, operation or failure to perform (whether consequential or otherwise) as determined by the information provided, is not transferred to any other party as a result of the verification.

1.5.3 Required Statement by Globe Performance Solutions and the Verification Entity

Globe Performance Solutions and the Verification Entity believe that the vendor's technology can achieve the performance claim set out in this Verification Report. This belief is based on independent analyses of information and declarations provided by the vendor and of independently generated data, using verification protocols authorized for the Canadian ETV Program. No additional bench or field tests were carried out to corroborate the data provided. The verification of performance is also based on a use of the technology in accordance with the specified operating conditions.

The Government of Canada, Globe Performance Solutions and the Verification Entity make no express or implied guarantee or warranty as to the performance of the vendor's technology. Nor do they guarantee or warrant the vendor's technology to be free from any defects in workmanship, or the integrity or safety of the technology as a whole, or it's compliance with such governmental codes, standards and regulations as may be applicable.

1.5.4 Required Statement of Audit by the Verification Entity

We have audited the information in support of the performance claim of [INSERT COMPANY NAME HERE] as vendor of the environmental equipment or process described in the Verification Report. We have audited in accordance with the procedures prescribed in the Canadian ETV Program General Verification Protocol, and have expressed our opinion in our Verification Report, dated MMM,DD,YYYY.

In our opinion, the information contained in the Verification Report presents fairly the performance achieved by the technology or process under review and is consistent with the original data and technical information from which it was derived.

2.0 Review of Application

2.1 Introduction

This Chapter provides a summary of the information provided by the applicant as included with a pre-screening application form and the formal application form submitted to Globe Performance Solutions and reviewed by [INSERT VERIFICATION ENTITY NAME HERE] for the Canadian ETV Program.

[INSERT PROPONENT INFORMATION HERE]. provided the following to support the claim: [PLEASE LIST]:

2.2 Applicant Organization

Name: [PROPONENT]
Company:
Address:
Tel:
Fax:
Email:

2.3 Review of Application

The technology and all information provided by the Applicant with the Formal Application, the formal application binder and all subsequent transmittals to the Verification Entity were reviewed. The results of this Application Review are summarized in the Application Review Checklist (Table 1).

Table 1: Application Review Checklist – Mandatory Information

Ref.	Criteria		Information Provided	
			Yes[1]	No
1.1	Signed Formal Application		☐	☐
1.2	Signed Declaration Regarding Codes & Standards submitted with signed formal application		☐	☐
1.3	Technology provides an environmental benefit.		☐	☐
1.4	A copy of "Claim to be Verified" for each performance claim to be verified included with the Formal Application.		☐	☐
1.5	Performance Claim composed in a way that satisfies "Criteria for Specifying Claims" :		☐	☐
	1.5.1	Include Technology name (and model number)	☐	☐
	1.5.2	Include application of the technology	☐	☐
	1.5.3	Include specific operating conditions during testing	☐	☐

[1] Provide written justification for yes or no information provided.

Ref.	Criteria		Information Provided	
			Yes[1]	No
	1.5.4	Does it meet minimum Canadian Standards/Guidelines *	☐	☐
	1.5.5	Does it specify the performance achievable by the technology	☐	☐
	1.5.6	Is it the performance measurable	☐	☐
1.6	Standard operating practices and a description of operating conditions for each individual performance claim specified.		☐	☐
1.7	The proponent has supplied significant references describing or supporting scientific and engineering principles of the technology. (see Chapter 4)		☐	☐
1.8	Two or more names and contact information of independent (no vested interest in the technology) experts, qualified (backgrounds of experts are needed) to discuss scientific and engineering principles on which the technology is based. These experts must be willing to be contacted by the VE.		☐	☐
1.9	Brief summary of significant human or environmental health and safety issues associated with the technology. (Note: this criterion complements but does not replace the obligation for the applicant to submit a duly signed "Declaration Regarding Codes and Standards")		☐	☐
1.10	Brief summary of training requirements needed for safe, effective operation of technology, and a list of available documents describing these requirements. (Note: this criterion complements but does not replace the obligation for the applicant to submit a duly signed "Declaration Regarding codes and standards")		☐	☐
1.11	Process flow diagram(s), design drawings, photographs, equipment specification sheets (including response parameters and operating conditions), and/or other information identifying the unit processes or specific operating steps in the technology. If feasible, a site visit to inspect the process should be part of the technology assessment.		☐	☐
1.12	**Supplemental materials** (optional) have been supplied which offer additional insight into the technology application integrity and performance, including one or more of :			
	A copy of patent(s) for the technology, patent pending or submitted.		☐	☐
	User manual(s).		☐	☐
	Maintenance manuals.		☐	☐
	Operator manuals.		☐	☐
	Quality assurance procedures.		☐	☐
	Sensor/monitor calibration program.		☐	☐
	Certification for ISO 9001, ISO 14000, or similar program.		☐	☐
	Material Safety Data Sheet (MSDS) information.		☐	☐

Ref.	Criteria	Information Provided	
		Yes[1]	No
	Workplace Hazardous Materials Information System (WHMIS) information.	☐	☐
	Health and Safety plan.	☐	☐
	Emergency response plan.	☐	☐
	Protective equipment identified.	☐	☐
	Technical brochures.	☐	☐
1.13	The applicant provided adequate documentation and data. There is sufficient information on the technology and performance claim for the performance claim verification. [If necessary, the VE should communicate with Globe Performance Solutions to request copies of the necessary documentation and required data that are available to support the claims.]	☐	☐

2.3.1 Application Review Checklist Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

3.0 Review of Technology

3.1 Technology Review Criteria

Table 2 must be completed for each environmental technology performance claim (or group of claims). If the Verification Entity considers some criteria especially important, or has other comments, this information must be documented and included with the assessment and report. Short comments may be included directly within the checklist text, and questions that are not applicable should be so noted.

Table 2 Technology Review Criteria Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
Technology Description			
2.1	Technology based on scientific and technical principles. (It will be necessary for the VE to read the key articles and citations listed in the Formal Application. It may also be necessary to contact the independent experts listed in the Formal Application to obtain additional information.)	☐	☐

[2] Provide written justification for no meets criteria.

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
2.2	Technology supported by peer review technical literature or references. (Peer review literature and texts must be supplied with the Formal Application as well as relevant regulations and standards that are pertinent to the performance claim)	☐	☐
2.3	Technology designed, manufactured, and/or operated reliably. (historical data from the applicant, not conforming to all data criteria, may be useful for the VE to review to assess the viability of the technology not for verification, but for insight purposes)[3]	☐	☐
2.4	Technology designed to provide an environmental benefit and not create an alternative environmental issue. (e.g. it does not create a more hazardous and or unmanaged byproduct and it does not result in the transfer of an environmental problem from one media to another media without appropriate management of the subsequent contaminated media)	☐	☐
2.5	Technology conforms to standards for health and safety of workers and the public.[4] The vendor must submit a signed "Declaration Regarding Codes & Standards", with the Formal Application. The role of the Verification Entity is to ensure this signed document is included with the information that is reviewed for the performance claim verification	☐	☐

[3] Also note The VE should use best judgment and apply standards relevant to the technology sector to generally assess whether the technology has been designed and manufactured in an acceptable fashion. A critical assessment of the materials / apparatus used in the technology is beyond the scope of the Canadian ETV Program. Any assessment of the integrity of the manufacture of technology components must be performed by personnel whose experience and expertise qualify them to undertake this activity. It is not the responsibility of the Verification Entity to assess the integrity of materials and substances used in the manufacture of the technology, other than to understand their use and implication on the performance of the technology.

It is the vendor's responsibility to ensure that applicable regulations and guidelines are satisfied with respect to application of the technology. The vendor must submit a signed "Declaration Regarding Codes & Standards", generally with the Formal Application. The role of the Verification Entity is to ensure this signed document is included with the information that is reviewed for the performance claim verification.

Claim verification by the Verification Entity does not represent any guarantee of the performance or safety of the equipment or process. The Verification Entity shall not be liable in any way in the event that the device or process fails to perform as advertised by the supplier or as expected by the consumer. The Verification Entity shall not be liable for any injury to person or property resulting from the use of the equipment or process.

[4] For the purposes of the Canadian ETV Program, the health and safety issue has been defined as a subjective criteria, requiring a value judgment on the part of the reviewer as to the integrity or reliability of any or all health and safety documentation provided by the applicant. As such, the Verification Entity cannot assume any liability in making a "Best Professional Judgment" assessment of the technology using

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
Environmental Standards			
2.6	Technology achieves federal, provincial, and/or municipal regulations or guidelines for management of contaminated and or treated soils, sediments, sludges, or other solid-phase materials.	☐	☐
2.7	Technology achieves federal, provincial, and/or municipal regulations or guidelines for all (contaminated and or treated) aqueous discharges as determined by the applicants information.	☐	☐
2.8	Technology achieves federal, provincial, and/or municipal regulations or guidelines for all (direct or indirect) air emissions. If the environmental technology results in the transfer of contaminants directly or indirectly to the atmosphere, then, where required, all regulations or guidelines (at any level of government) relating to the management of air emissions must be satisfied by the applicant's information.	☐	☐
Commercial Readiness			
2.9	Technology and all components (apparatus, processes, products) is full-scale, commercially-available, **or alternatively see 2.10 or 2.11**, and, data supplied to the Verification Entity is from the use or demonstration of a commercial unit.	☐	☐

these criteria.

(continued footnote 5 from Ref. 2.5) A critical validation of the Health and Safety aspects of the vendor's technology is beyond the scope of the Canadian ETV Program. Any validation of health and safety issues must be performed by personnel whose experience and expertise qualify them to undertake these activities. Staff from noted organizations and agencies [e.g., Health and Welfare Canada (H&W), Provincial Labour Ministries, Industrial Accident Prevention Association (IAPA), [US] Occupational Safety and Health Association (OSHA), water pollution control agencies, province/state health departments, fire protection associations, etc.], may be able to provide advice or technical services on these issues. It is **NOT** the responsibility of the Verification Entity to validate the Health and Safety aspects of the technology.

It is the vendor's responsibility to ensure that regulations and guidelines are satisfied in the application of the technology. The Verification Entity can request additional written confirmation from the applicant that the company has sufficient documentation to address worker health and safety issues and requirements related to the use of the technology, including an Emergency Response Plan.

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
2.10	Technology is a final prototype design prior to manufacture or supply of commercial units, **or alternatively see 2.11**, Note: Verification of the performance claim for the technology is valid if based on a prototype unit, if that prototype is the final design and represents a pre-commercial unit. The verification will apply to any subsequent commercial unit that is based on the prototype unit design. The verification will not be valid for any commercial unit that includes any technology design change from the prototype unit used to generate the supporting data for the verification.	☐	☐
2.11	Technology is a pilot scale unit used to provide data which when used with demonstrated scale up factors, proves that the commercial unit satisfies the performance claim.[5]	☐	☐
Operating Conditions			

[5] In exceptional situations, data from a pilot scale unit may be used to validate a performance claim. This situation can be permitted if the pilot scale unit is a "scaled down" model of a full size commercial unit and engineering scale-up factors have been provided by the applicant as part of the verification process. The performance claim verification must include validating the scale-up factors.

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
2.12	All operating conditions affecting technology performance and the performance claim have been identified.	☐	☐
2.13	The relationships among operating conditions and their impacts on technology performance have been identified. Note: It is the responsibility of the VE to understand the relationship between the operating conditions and the performance of the technology, and to ensure that the impacts of the operating conditions and the responses of the technology are compatible.	☐	☐
2.14	Technology designed to respond predictably when operated at normal conditions (i.e. conditions given in 2.12), **and/or alternatively see 2.15,** Note: The Verification Entity must be satisfied that these data do not demonstrate a performance that is different than the performance indicated in the Performance Claim to be validated.	☐	☐
2.15	Effects of variable operating conditions, including start up and shut down, are important to the performance of the technology and have been described completely as a qualifier to the performance claim under assessment.	☐	☐
Throughput Parameters			
2.16	Effects of variable contaminant loading or throughput rate must be assessed and input/output limits established for the technology. Note: If the application of the technology is to a variable waste source or expected (designed) variable operating conditions, then it will be necessary to establish acceptable upper and lower ranges for the operating conditions, applications and/or technology responses. Sufficient, quality data must be supplied to validate the performance of the technology at the upper and lower ranges for the operating conditions, applications and or technology responses detailed in the performance claim.	☐	☐

Other Relevant Parameters/Variables/Operating Conditions

The Verification Entity is expected to understand the technology and identify and record **all** **relevant** criteria, parameters, variables or operating conditions that potentially can or will affect the performance of the technology under assessment. It is practical to include all of these variables in Table 2 (from 2.17 to …).

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
2.17		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐
2…	Continue on attached page(s) as required.	☐	☐

[2] Provide written justification for yes or no meets criteria.

3.1.1 Technology Review Checklist Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

4.0 Review of Test Plan, Test Execution and Data

4.1 Review of Test Plan and Execution of Test Plan

Table 3 Verification Study Design Assessment Criteria Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[6]
3.1	Was a statistician, or an expert with specialized capabilities in the design of experiments, consulted prior to the completion of the test program, and if so please provide the contact details.[7]	☐	☐
3.2	Is a statistically testable hypothesis or hypotheses provided? (so that an objective, specific test is possible)[8]	☐	☐
3.3a-c	Does the verification study generate data suitable for testing the hypothesis being postulated?[9] Namely:	☐	☐
3.3a	Does the study measure the parameters used in the performance claim hypothesis?	☐	☐
3.3b	Does the study control for extraneous variability?	☐	☐
3.3c	Does the study include only those effects attributable to the environmental technology being evaluated?	☐	☐
3.4	Does the verification study generate data suitable for analysis using the SAWs? (i.e. it is preferable that tests are designed with the SAWS in mind before test plans are written)	☐	☐
3.5	Does the verification study generate data suitable for analysis using other generic experimental designs (ANOVA etc)? (clearly, verification studies should be designed with the final data analysis in mind to facilitate interpretation and reduce costs)	☐	☐

[6] Provide written justification for yes or no meets criteria.
[7] An expert statistician can help determine during the experimental design which experimental variables need to be controlled and or monitored so as to be able to defend a verification claim
[8] The hypothesis that Statistical Analysis Worksheets will test are of the general form:
What is the degree of confidence around a measured result?
Is a mean equal to a specified value?
Is a median equal to a specific value?
Is mean 1 = mean 2 ?
Is median 1 = median 2 ?
Is variance 1 = variance 2 ?
Can a process change an influent/product/waste by 'p' percent?
Are two paired measurements different?
[9] Note: When data are not available on a specific parameter, it may be possible to use data on a surrogate parameter that has known correlation to the unmeasured parameter. In this case, the correlation must be clearly defined, demonstrated and based on sound scientific, engineering and or mathematical principles. The applicant must submit that data for their set of tests.

3.6	Are the appropriate parameters, specific to the technology and performance claim, measured? (it is essential that the VE and the technology developer ensure that all parameters – e.g. temperature etc - that could affect the performance evaluation are either restricted to pre-specified operating conditions or are measured)	☐	☐
3.7a-d	Are samples representative of process characteristics at specified locations?. **namely:**		
3.7a	Are samples collected in a manner that they are representative of typical process characteristics at the sampling locations for example the samples are collected from the source stream fully mixed etc	☐	☐
3.7b	Is data representative of the current technology?	☐	☐
3.7c	Have samples been collected after a sufficient period of time for the process to stabilize?	☐	☐
3.7d	Have samples been collected over a sufficient period of time to ensure that the samples are representative of process performance?	☐	☐
3.8	Are samples representative of operating conditions? Note: A time lag occurs between establishing steady state conditions and stabilization of the observed process performance. This time lag depends in part on the time scale of the process. (i.e. for a Completely Stirred Tank Reactor (CSTR) flow-through system, the time scale is determined by the residence time of the contaminants in the reactor. It is usual that at least three residence times are required to achieve effective stabilization. Therefore if sampling has been performed from a CSTR, then sampling should have only begun after at least three hydraulic residence times had occurred, and testing continued for at least an additional three residence times to ensure that the aggregate data set is representative of process performance)	☐	☐
3.9	Are samples representative of known, measured and appropriate operating conditions? (Note: this includes technologies that operate on short cycles and so have start and stop cycles which affects the operation of the technology). If the operating conditions are not vital but are recommended, then the reviewer must evaluate operating conditions,	☐	☐
3.10	Were samples and data prepared or provided by a third party? (Note: In some cases, where the expertise rests with the applicant, an independent unbiased third party should witness and audit the collection of information and data about the technology. The witness auditor must not have any vested interest in the technology.)	☐	☐

3.11a-c	**Verification Study Design is Acceptable Namely:**		
3.11a	**The samples have been collected when the technology was operated under controlled and monitored conditions.**	☐	☐
3.11b	**A verification study design should have been established prior to the test to ensure that the data were collected using a systematic and rational approach**	☐	☐
3.11c	**Verification Study Design should have defined the acceptable values or ranges of values for key operating conditions, and the data collection and analysis methodology**	☐	☐

4.1.1 Verification Study Design Assessment Criteria Checklist Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

4.2 Data Validity Checklist

The data validity checklist criteria help the VE determine whether a datum represents the conditions described in the performance claim. The data validity checklist also ascertains whether or not samples have been collected, transported and analyzed in a manner that does not introduce undue extraneous variability.

Table 4 A Data Validity Checklist

Ref.	Criteria	Meets Criteria	
		Yes	**No**[10]
4.1	Were appropriate sample collection methods used (e.g. random, judgmental, systematic etc?). For example: simple grab samples are appropriate if the process characteristics at a sampling location remain constant over time. Composites of aliquots instead may be suitable for flows with fluctuating process characteristics at a sampling location. Note: Sampling methods appropriate for specific processes may sometimes be described in federal, provincial or local monitoring regulations	☐	☐
4.2	Were apparatus and/or facilities for the test(s) adequate for generation of relevant data? (i.e. testing was performed at a location and under operating conditions and environmental conditions for which the performance claim has been defined.)	☐	☐
4.3	Were operating conditions during the test monitored and documented and provided?	☐	☐
4.4	Has the information and or data on operating conditions and measuring equipment measurements and calibrations been supplied to the Verification Entity?	☐	☐
4.5	Were acceptable protocols used for sample collection, preservation and transport (acceptable protocols include those developed by a recognized authority in environmental testing such as a provincial regulatory body, ASTM, USEPA, Standard Methods)?	☐	☐

[10] Provide written explanations for yes or no meets criteria.

4.6	Were Quality Assurance/Quality Control (QA/QC) (e.g. use of field blanks, standards, replicates, spikes etc) procedures followed during sample collection? A formal QA/QC program, although highly desirable, is not essential, if it has been demonstrated by the vendor's information that quality assurance has been applied to the data generation and collection.	☐	☐
4.7	Were samples analyzed using approved analytical protocols? (e.g. samples analyzed using a protocol recognized by an authority in environmental testing such as Standard Methods, EPA. ASTM etc. Were the chemical analyses at the site in conformance with the SOPs (Standard Operating Procedures) ?	☐	☐
4.8	Were samples analysed within recommended analysis times (especially for time sensitive analysis such as bacteria)	☐	☐
4.9 a-e	Were QA/QC procedures followed during sample analysis Including?	☐	☐
4.9a	Maintaining control charts	☐	☐
4.9b	Establishing minimum detection limits,	☐	☐
4.9c	Establishing recovery values	☐	☐
4.9d	Determining precision for analytical results	☐	☐
4.9e	Determining accuracy for analytical results	☐	☐
4.10 a-c	Was a chain-of-custody (full tracing of the sample from collection to analysis) methodology used for sample handling and analysis. Namely:	☐	☐
4.10a	Are completed and signed chain-of-custody forms used for **each sample** submitted from the field to the analytical lab provided for inspection to the Verification Entity?	☐	☐
4.10b	Are completed and easily readable field logbooks available for the VE to inspect?	☐	☐
4.10c	Are their other chain-of-custody methodology actions and documentation recorded/available (e.g. sample labels, sample seals, sample submission sheet, sample receipt log and assignment for analysis)	☐	☐
4.11	**Experimental Data Set is Acceptable** **(the quality of the data submitted is established using the best professional judgment of the VE)**	☐	☐

4.2.1 Data Validity Checklist Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

4.3 Remote sensing data (e.g. telemetry)

For Data produced at sites which need the data to be sent electronically from the on-site instrument to the data receiving site (node) needs to be assessed for its integrity. Table 4b details the areas of the data integrity needing assessment by the VE.

Table 4b Remote sensing data

Ref.	Criteria	Meets Criteria	
		Yes	**No**[11]
4.12	Does the remote data device (e.g. meter) authenticate all sending and receiving nodes prior to any data transfer	☐	☐
4.13	Is data sent from remote monitoring device encrypted during transfer	☐	☐
4.14	Is the data received with 100% integrity	☐	☐
4.15	What methods can be demonstrated that the data is received with 100% accuracy	☐	☐
4.16	**Experimental Remote Data Set best practices are Acceptable (the quality of the data submitted is established using the best professional judgment of the VE)**	☐	☐

4.3.1 Remote Sensing Data Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

Note, depending on the nature of the verification, the checklists may need to be modified and or new questions developed based on discussion with Globe Performance Solutions and the vendor.

4.3.2 Other Verification Topics Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

4.4 Data Analysis Checklist

The intent of the data analysis checklist is to ensure that the appropriate statistical tools can be used in a rigorous, defensible manner.

[11] Provide written explanations for yes or no meets criteria.

Table 5 Data Analysis Checklist

Ref.	Criteria	Meets Criteria	
		Yes	No[2]
5.1	Does the analysis test the performance claim being postulated? (When conducting performance evaluations, under the Canadian ETV Program, the alternative hypothesis of a "significant difference" without stating the direction of the expected difference will usually be unacceptable)	☐	☐
5.2	Does the analysis fit into a generic verification study design? (Many other "generic" designs exist that are not explicitly covered by the Canadian ETV Program (e.g. ANOVA, ANCOVA, regression etc) that are potentially useful)[12]	☐	☐
5.2 a-c	Are the assumptions of the analysis met. Namely: (a negative response to 3.30 a-c means the VE needs to request further information)		
5.2.a	Did the data analyst check the assumptions of the statistical test used?	☐	☐
5.2.b	Are the tests of assumptions presented?	☐	☐
5.2.c	Do the tests of the assumptions validate the use of the test and hence the validity of the inferences?	☐	☐
5.3	**Data Analysis is Acceptable** The data analysis is acceptable if the statistical test employed tests the hypothesis being postulated by the technology developer, the assumptions of the statistical test is met and the test is performed correctly.	☐	☐

4.4.1 Data Analysis Checklist Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

4.5 Data Interpretation Checklist

The intent of the data interpretation checklist is to ensure that the data analyses results are interpreted in a rigorous, defensible manner. The checklist also emphasizes that an initial performance claim may be rewritten and updated to better reflect what the data support, using the expertise of the VE and other pertinent resources.

Table 6 Data Interpretation Checklist

[12] Examples of potentially useful verification study designs or analyses not covered by the Canadian ETV Program are:

- completely randomized designs with more than two treatments (ANOVA);
- designs where some of the operating conditions vary widely enough to require acknowledgement both in the experimental design and analysis stage. (ANCOVA, regression);
- analysis of count data such as microbial counts; and,
- analysis of proportional data such as proportion of organisms responding to a treatment.

Ref.	Criteria	Meets Criteria	
		Yes	**No[13]**
6.1a	Are the results statistically[14] or operationally significant? Did the verification result in a statistically significant test of hypothesis?	☐	☐
6.1b	To be operationally significant, does the technology meet regulatory guidelines and applicable laws?	☐	☐
6.2	Does the verification study have sufficient power to support the claim being made? *Note: For verification study designs where acceptance of the null hypothesis results in a performance claim being met, the statistical power of the verification study must be determined A statistical power of at least 0.8 is the target. If the power of the verification experiment is less than this value the VE should contact Globe Performance Solutions to discuss an appropriate course of action.* *See Appendix A for examples on calculating sample size*	☐	☐
6.3	Is the interpretation phrased in a defensible manner? Note: The final performance claim should reflect any changes to the claim made during the course of the analyses, variations or restrictions on operating conditions, etc. that changed the scope of the performance claim. The initial performance claim should be viewed as a tentative claim that is subject to modification as the verification progresses. A thoughtful open-minded verification will in the end, prove to be of greatest benefit to the technology developer.	☐	☐
6.4	**Data Interpretation is Acceptable** The data interpretation is acceptable if the data analyses results are reviewed in a manner that emphasizes the applicability to the specific performance claim and the statistical power of the verification experiment.	☐	☐

4.5.1 Data Interpretation Checklist Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION HERE]

[13] Provide written justification for yes or no meets criteria.

[14] In some cases, a new statistical approach may be necessary in order to analyze the data provided. If the existing Statistical Analysis Worksheets (SAWs) provided in the General Verification Protocol (GVP) do not apply, any other proposed approaches should be discussed with and approved by Globe Performance Solutions. In these cases, the preferred course would be to have additional SAWs developed by Globe Performance Solutions.

5.0 Statistical Analysis of the Performance Claim(s)

This step in the verification process involves the statistical analysis review of the performance claim(s). The Statistical Analysis Worksheets (SAWs) contained in Appendix A may be used to mathematically evaluate the performance claim(s).

5.1 Performance Claim

Title of Performance Claim: _____

5.2 Performance Claim(s) Verification

The verification of each technology performance claim(s) requires application of the Statistical Analysis Worksheets (SAWs) to all data sets that were rated as satisfactory from the data assessment process.

The data set(s) provided to support the performance claim should be evaluated using the Statistical Analysis Worksheets in Appendix A. The SAWs were chosen to provide analytical methods for the most common types of data sets generated by verification experiments. They are suitable for use by the non-statistician, provided test assumptions are verified and the concepts emphasized in the GVP and SAWs are understood and used when data interpretations are made. Use of the SAWs is described in the Case Studies in Appendices E and F.

A summary of the Statistical Analysis, highlighting the data sets and the specific statistical analysis worksheets used, can be summarized in the Table 7 below.

Table 7 Summary of Acceptable Data Sets for Performance Claim Verification

Acceptable Data Set(s) Identification	SAWs Used[15]	Support Claim	
		Yes	No
		☐	☐
		☐	☐
		☐	☐
		☐	☐
		☐	☐

[15] Refer to Appendix A

6.0 Establishment of the Audit Trail

- Summary of key supporting documents

As a summary of some of the most important paper documents that the VE needs to possess, refer to Table 8.

Table 8 Key documents

KEY DOCUMENTS	Present	Absent
Raw data sheets and summary data	☐	☐
Signature pages	☐	☐
Signed Formal Application	☐	☐
Declaration Regarding Codes & Standards	☐	☐
Patent(s)	☐	☐
Sample security: e.g. chain of custody sheets for each sample *	☐	☐
Operation and maintenance manual	☐	☐
Field notebooks	☐	☐
Certificate of accreditation of laboratories	☐	☐

* These items may or may not be available for the Verification Entity but are useful in determining reasons for data discrepancies etc. Where applicable and depending on the nature of the verification test program the VE should request to see these asterisked items.

6.1 Audit Trail Comments

[VERIFICATION ENTITY INSERT COMMENTS RELATING TO EACH SECTION OF THE AUDIT TRAIL HERE]

7.0 Conclusions

[Within this section, The VERIFIED performance claim should be given in the box in the verification report (e.g. as below.)]

Verified Performance Claim:

[Additionally, overall comments and conclusions about the verification need to be described here by the Verification Entity]

Appendices to the Verification Report

Appendix A Statistical Analysis Worksheets
Appendix B Selected Statistical tables
Appendix C Selected References
Appendix D Append Critical material as necessary

FIGURES

LIST FIGURES AS NECESSARY

TABLES

LIST TABLES AS NECESSARY INCLUDING THE FOLLOWING:

TABLE 1: PPLICATION REVIEW CHECKLIST – MANDATORY INFORMATION.
Table 2: Data Generation Review
TABLE 3: TECHNOLOGY REVIEW CRITERIA CHECKLIST.
Table 4: Verification Study Design Assessment Criteria Checklist.
TABLE 4A: DATA VALIDITY CHECKLIST.
TABLE 4B: REMOTE SENSING DATA
TABLE 5: DATA ANALYSIS CHECKLIST.
TABLE 6: DATA INTERPRETATION CHECKLIST.
TABLE 7: SUMMARY OF ACCEPTABLE DATA SETS FOR PERFORMANCE CLAIM VERIFICATION.
TABLE 8: KEY DOCUMENTS