MILLENNIUM
EMS Solutions Ltd.

#148, 2257 Premier Way
Sherwood Park, AB T8H 2M8
tel: 780.496.9048
fax: 780.496.9049

Suite 325, 1925 18 Avenue NE
Calgary, AB T2E 7T8
tel: 403.592.6180
fax: 403.283.2647

#102, 11312 98 Avenue
Grande Prairie, AB T8V 8H4
tel: 780.357.5500
fax: 780.357.5501

toll free: 888.722.2563
www.mems.ca

Regional Background Soil Quality Assessment Methodology Report

17-SGRC-05

Prepared for:
Petroleum Technology Alliance Canada

Prepared by:
Millennium EMS Solutions Ltd.
Suite 325, 1925 18 Avenue NE
Calgary, Alberta
T2E 7T8

March 2019
File #17-00395

Table of Contents

List of Tables

List of Appendices

1.0 INTRODUCTION

Millennium EMS Solutions Ltd. (MEMS) was retained by Petroleum Technology Alliance Canada (PTAC) to develop an approach for regional assessment of background soil quality. The work was funded under the Alberta Upstream Petroleum Remediation Fund (AUPRF), under agreement 17-SGRC-05.

1.1 Background

Consideration of background salinity is an important component of the assessment of sites with potential for salinity issues; understanding of background soil chemistry can also be important for other substances as well, such as organic substances which may be of natural origin (*e.g.*, polycyclic aromatic hydrocarbons) and metals. Adequate characterization of background soil chemistry can require a relatively large number of samples, particularly in areas where there is high variability in background concentrations. In the absence of sufficient background data, sites can be identified as contaminated when soil quality is within the actual background range, often leading to unnecessary or non-beneficial remediation.

Previous attempts have been made to develop regional background salinity databases in effort to evaluate background conditions with less site-specific data. However, in the absence of clearly defined regulatory guidance specifying the requirements for this approach, each effort is evaluated individually and without specific evaluation criteria. As a result, proponents do not know what is required, and regulators are not able to perform a timely and consistent review.

The intent of the current report is to outline a scientifically defensible outcome-based approach to regional background salinity assessment. The approach includes guidance on when a regional assessment can be considered, minimum requirements for the assessment, and how to determine whether a specific site can be evaluated using the regional background data set. The work draws on existing and under development regional assessments as well as industry-provided data sets from areas expected to be good candidates for regional assessment. It also provides a foundation by which other parameters may be assessed.

1.2 Objectives

The overall objective of the project is to develop a defensible approach for assessing background salinity on a regional basis, to reduce the costs of detailed background soil quality assessments and reduce unnecessary remediation. To meet this objective the approach looks to:

- define a consistent procedure for use across a defined Study Area that is acceptable to Alberta Environment and Parks (AEP) and the Alberta Energy Regulator (AER);

- be used to minimize assessment and remediation costs to the extent practicable by sharing data; and
- reduce unnecessary remediation on low risk sites.

2.0 KEY PROGRAM ELEMENTS

In order to conduct a regional background soil quality assessment, the following key steps are required:

1. **Scoping** – an understanding of when background samples are required and how best to acquire those samples.
2. **Defined Study Area** - the Study Area is based not only on geographical location but on set criteria that ensure that concentrations can be meaningfully compared between background sampling locations and investigation sites. The information collected must include enough detail on sampling location (*e.g.*, gradient, soil type, location) to determine whether a regional background concentration data set is applicable.
3. **Data Quality Assessment (DQA)** - regional background data sets must adequately represent the Study Area and be unaffected by anthropogenic input(s). Typically, data sets will be separated by criteria including depth/stratigraphic interval. Key components of a DQA include planning site data collection and laboratory submission.
4. **Statistical Evaluation** - statistical comparison of investigation site data to regional background data to determine the extent of anthropogenic inputs above background concentrations.

Each of these steps is detailed in the sections below. An example statistical walk through of a Site assessment has been provided in Appendix A.

3.0 SCOPING

The first step in conducting a regional background soil quality assessment is collecting and evaluating all available information that is relevant to the investigation site and background soil chemistry. Scoping the relevant information will help determine if additional background sampling is required and how best to acquire those samples.

The adequacy and sufficiency of existing information may be evaluated based on a variety of factors, including sample quantity, location, quality, and data gaps. In scoping adequate background locations (establishing the Defined Study Area), samples must be representative of sample locations as described in Section 4.0. Additionally, a sufficient number of samples is required for the purposes of performing statistical evaluation, and to perform statistical evaluations with the preferred level of statistical power (described in Section 6.0), consideration for the level of effort in obtaining such samples is an important consideration at the Scoping stage. Finally, all existing data for the investigation site should also be reviewed for quality and completeness (US EPA, 2002), and a decision should be made on whether a background comparison is warranted; background assessment may not be warranted if the investigation site data provides clear evidence of anthropogenic input.

4.0 DEFINED STUDY AREA

The defined Study Area is the location where background samples will be collected for comparison with investigation sites. The Study Area is based not only on geographical location but on set criteria that ensure that concentrations can be meaningfully compared. A discussion of the requirements for defining a Study Area for the development of a regional background concentration data set is described below.

4.1 Regional Background Concentration Data set(s)

4.1.1 Study Area Definition

In order to assess regional background soil quality, the Study Area must be clearly defined. At a fundamental level, the key component of this definition is that the soil would be basically the same, particularly with respect to the substance being assessed (*e.g.,* background salinity). Therefore, the Study Area needs to consider factors which affect natural or background salinity. These include (from Alberta Agriculture and Forestry, 2004; Chang *et al*, 1985):

- the parent material of the soil;
- soil texture;
- groundwater depth;
- climate;
- slope position, or more specifically the presence of groundwater discharge zones; and
- surface water features, particularly sloughs.

Therefore, in effort to establish a Study Area, it must be limited to an area with similar climate and similar soils. Furthermore, within the Study Area, specific sampling locations would need to consider soil texture, slope position and distance from surface water for a meaningful comparison of results.

4.1.2 Defining Study Area Boundaries

To define the Study Area boundaries, the *Canada – Alberta Environmentally Sustainable Agricultural* (CAESA) soil inventory project (SIP) can be used (CAESA, 1998). The CAESA SIP employs a hierarchical ecological land classification (ELC) system stemming from *A National Ecological Framework for Canada* (Ecological Stratification Working Group, 1995), analogous to, or a predecessor of, the *Natural Regions and Subregions of Alberta* (Natural Regions Committee, 2006). At the most refined level, the CAESA SIP is commonly used to identify specific soil polygon units and soil attributes, viewed through the *Agricultural Regions of Alberta Soil Inventory Database* (AGRASID). The hierarchical ELC is as follows:

- Ecozone (analogous to a Natural Region);
- Ecoregion (analogous to a Natural Subregion);
- Ecodistrict or Land Resource Areas (LRA; subset of an Ecoregion);
- Land System (subset of an Ecodistrict); and
- Soil Landscape Inventory (subset of a Land System; AGRASID Soil Polygon).

An ecodistrict is proposed as the base unit for initial determination of the Study Area boundaries, with a further review of the corresponding subset of land systems to define the final boundaries. The land systems within an ecodistrict are expected to be generally consistent in relation to soils, topography, landscape and surface geology.

An ecodistrict is defined as a "part of an ecoregion characterized by distinctive assemblages of relief, geology, soils, vegetation, water, fauna and land use"(Ecological Stratification Working Group, 1995). In Alberta there are 136 ecodisticts, ranging in size from approximately 10 to 50 townships (93,250 to 466,250 hectares). Within an ecodistrict there are several land systems. Land systems are differentiated based on topography, surficial geology, and soils, to a more refined level than the ecodistrict. Within Alberta, land systems are as small as 1 township (9,325 ha), with an average size of approximately 3 townships (27,975 ha) (Government of Alberta, 2016). The land systems within an ecodistrict requires review based on land morphology, soil order and major soil type; land systems that differ from the majority within the ecodistrict would then be removed from the Study Area.

Overall, the initial Study Area would therefore comprise one or more land systems with similar land morphology, soil order and major soil type within a single ecodistrict. Following the establishment of the ecodistrict the ecoregion should also be defined for the Study Area; including information on evaporation and precipitation (which affect salt movement direction), water bodies, topography, vegetation, geology, hydrogeology and soil types.

4.1.3 Study Area Characterization

4.1.3.1 Sampling Methodology

Regional background soil samples must be free of effects from anthropogenic point source contamination. For background salinity, a detailed salinity analysis is necessary to demonstrate the cause of any elevated electrical conductivity (EC). Background data samples can include samples collected from unimpacted locations at an investigation site (*e.g.*, background samples collected during a Phase 2 investigation) or samples collected away from specific sites to better characterize the Study Area as a whole.

Sampling must be conducted in accordance with AEP/AER policy and the Guidance Manual for Environmental Site Characterization in Support of Environmental and Human Health Risk Assessment (CCME, 2016).

In order to characterize background soil quality on a regional scale, sufficient information needs to be collected from the sampling location to confirm that it reflects background and is similar to the locations it will be compared against. Key information includes:

- topography (slope and elevation);
- land use;
- depth of sample;
- depth to groundwater; and
- soil type.

In order for regional guidelines to be applied, the above key parameters need to be similar in all background samples from the Study Area.

4.1.3.2 Topography (Slope and Elevation)

Soil samples must be collected from an area of similar slope and elevation as the shape and elevation of the land surface influences the redistribution of water received as precipitation. In general, low lying depressional areas receive water whereas elevated sloping surface lose water *via* runoff. Ultimately, the net result is drier less developed soils on sloping surfaces and deeper more strongly developed soil profiles in moist lower regions.

4.1.3.3 Land Use

Select land use practices contribute to soil salinization by affecting both the quantity and the flow of water and salts through the root zone (Wiebe *et al.*, 2005). For example, the practice of summerfallow may act to increase water content in the root zone which may elevate the water table and increase the level(s) of soluble salts at or near surface. In contrast, permanent and continuous cropping each promote water use from the root zone, lower the water table and draw salinity downward out of the rooting zone.

4.1.3.4 Depth of Soil Sample

The biologically active zone of soil is the region where microbial and invertebrate activity, which influences the overall ecological health occurs and is considered to be dictated by plant rooting activity. The ability to delineate the biologically active zone at a contaminated site is a prerequisite for describing exposure for ecological resources, potential transport mechanisms and quantifying associated risk (Sample *et al*., 2014). Within rooting zones, root associated microbes, including endophytes, interact and mediate important physiological processes, especially nutrient acquisition and plant fitness to abiotic stresses. Contaminants in soil below the biologically active zone are not in direct contact with receptors, do not represent a complete ecological exposure pathway and thus do not pose a direct threat to terrestrial plants or animals. Therefore, if the purpose of comparison is to establish background salinity effect on plant health, it would be inappropriate to make such comparisons with soils collected below typical plant rooting depths.

In Alberta, there are several guidance documents that specifically reference the depth of soil samples as a condition of their application:

- Alberta Environment and Parks (AEP) *Tier 1 and Tier 2 Soil and Groundwater Guidelines* (2016a, b):
  - Applies to all land uses and governs by depth;
  - The ecological direct contact pathway may be eliminated at depths exceeding 3.0 meters, if an alternative guideline is available (*i.e.,* management limit), which currently applies only to petroleum hydrocarbon (PHC) fractions F1 to F4.
- *Salt Contamination Assessment and Remediation Guidelines* (SCARG) (AENV 2001):
  - Applies to all land uses and governs by lithology; and
  - The ecological direct contact pathway has separate guidelines derived specifically for surface soil (defined as the A-horizon) and subsoil (defined as the B- and C-horizons and the upper portion of the parent material).
- Contaminated Sites Management: Subsoil Salinity Tool (SST) (ESRD 2014):
  - Applies to all land uses and governs by depth;
  - The rooting zone of the ecological direct contact pathway exists to a depth of 1.5 meters;
  - Applicable only to chloride at depth greater than 1.5 m; and
  - Used to predict upward migration of chloride into root zone (*i.e.,* 1.0 to 1.5 m) and calculates a predicted EC value within root zone and compares to SCARG.

Therefore, soil samples collected for the purpose of establishing background conditions under a background soil quality program must incorporate the regulatory regime in which they are to be applied. As an example, soil samples collected from 0 to 1.5 mbgs may have broad applicability as it pertains to the comparison of one surface soil to another, but under SCARG, the absence of a defined horizon would not allow for meaningful comparison.

4.1.3.5 Depth to Groundwater

Movement and translocation of aqueous solutions through the unsaturated zone is affected by deep drainage and recharge rates. Understanding the physical parameters that define deep drainage is therefore critical for the assessment of risk associated with anthropogenic activities, including applications of herbicides, fertilizers, and uncontrolled releases of industrial fluids. Deep drainage has often been discussed as a function of climatic and geomorphic variability (Gee *et al*., 2005; Vaccaro, 1992). Regions of relatively high precipitation have been routinely assessed and methods of determining deep drainage in humid climates are well developed (Delin and Risser, 2007; Gee and Hillel, 1988).

Transport of water and solute through the unsaturated zone can be separated into two individual processes. The first is movement of water and solute into and through the root zone. The fluxes associated with inter-root zone transport are transient and include infiltration and evapotranspiration, both of which result in concentration and hydraulic gradients. The second process involves the movement of water and solute below the root zone, resulting in ground water recharge (Gee and Hillel, 1988; Dyck, 2001). The advective movement of water below the root zone is thought only to occur as a result of high precipitation events that exceed the rate of evapotranspiration (Dyck *et al*., 2003; Gee and Hillel, 1988).

In simple terms, it is important to know both the depth to groundwater as well as the precipitation norms for the region, when establishing whether or not soils from background locations are suitable for direct comparison. As an example, the comparison of two soils, one being from an arid deep draining location with a low water table and the other from a moist, poorly drained location with a high water table, would not be suggested, albeit both samples may have been collected from the same depth.

4.1.3.6 Soil Type

Using the knowledge of soil formation, the applicability of background sites can be assessed by evaluating the soils that are present and inferring as to the soil forming factors that influenced their creation. This information can provide a method for establishing whether or not an upward or downward migration of COPCs is suspected at the regional scale, and whether or not the soil formation process in two regions are distinctly dissimilar.

For example, the textural and chemical properties of the parent material largely determine the state and abundance of minerals in the soil (Wilson, 2006; Ayres *et al*., 1985). With respect to chemical composition, the majority of the till material deposited in the Prairie Provinces was of calcareous nature with high concentrations of organic sulphur, resulting from the stripping of Cretaceous shales, limestone and dolostone (Schreiner, 1990). Oxidation of organic sulphur found in the till material resulted in an increase in sulphate salts concentrations. The distribution of sulphates within the soil profile at the time of deposition was relatively uniform; however, over time water infiltration resulted in redistribution of sulphate within the soil profile (Woods, 2006). Depletion of sulphate from the upper soil profile indicates long-term downward movement of water resulting in deep drainage. This downward migration is an indication that the potential exists for COPCs to migrate out of the rooting zone over time. It is also an indication that soils with different pedogenic salt distributions are from dissimilar soil formation processes.

4.2 Compilation of Background Data Set(s)

In select cases it may be appropriate to conduct additional sampling and develop multiple background data sets for a Study Area. This may be necessary to effectively characterize background soil quality on a regional scale, particularly if the Study Area exhibits a range of the key parameters identified in Section 4.1.3.

Examples of cases where it may be appropriate to develop multiple data sets for single Study Area include:

- There is more than one common soil type present within the Study Area; in this case a separate background data set may be generated for each common soil type.
- There are vertical trends in salinity – *e.g.,* a saline layer at a specific depth over a large area; in this case it would be appropriate to separate out the saline layer as a separate data set.
- The depth to groundwater is variable across the Study Area; in this case it would be appropriate to separate background samples where there may be an influence from groundwater from those collected at locations with deep groundwater.

In such cases, the same requirement for the establishment of a background soil quality value would be required, but instead of a single background data set, the minimum data requirements would need to be met for each of the data set(s) being developed.

5.0 DATA QUALITY ASSESSMENT

Data quality assessment (DQA) plans should be used to ensure that background data collection meets pre-determined technical and quality objectives of the project. Regional background data sets must adequately represent the Study Area and be unaffected by anthropogenic inputs. Key components of a DQA include planning site data collection and laboratory submission.

A general outline of a DQA plan is discussed in the sections below. Additional resources for information about DQA plans include the Interstate Technology and Regulatory Council's (ITRC) *Groundwater Statistics and Monitoring Compliance: Statistical Tools for the Project Life Cycle* (2013) and US

Environmental Protection Agency's (US EPA) *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites* (2002) guidance documents.

5.1 Investigation Site Data Collection

There are additional uncertainties to consider when collecting soil data from the investigation site as background. Namely, there is an increased potential for unintended sampling of soils that have been impacted or disturbed by anthropogenic activity. Where possible, soil samples should not be collected downgradient or in the direction of groundwater flow from active or historical infrastructure.

The following provides a general guidance for background sampling within an investigation site:

- Samples should not be collected from fill, especially in the case of non-native fill;
- Samples should not be collected from native soils that have been overlain by fill. For example, an A-horizon that has been overlain during a pad construction would not be considered a suitable background sample for comparison to other A-horizons that remain at soil surface;
- Samples should not be collected within 10 m of historical or active infrastructure nor in areas where current activity is ongoing (*e.g.*, storage areas/laydown yards, truck turn outs, *etc.*); and
- Caution should be given to sampling in areas designated for pond water release or snow dump.

Planning for a background sampling location within a site boundary requires an understanding of both historical and current operations. An understanding of the sites spill history will also prove beneficial in the determination of suitable background sampling locations (avoid sampling in areas that may have been impacted by historical release).

5.1.1 Exclusion of Data

Data points are removed from the background data set if:

- there is indication of anthropogenic contamination;
- the sample was collected in the immediate vicinity of surface water, including seasonal sloughs;
- the sample was collected from a location where slope changes or at a groundwater recharge or spring location; and
- the soil type of the sample is not consistent with the other samples in the data set.

5.2 Laboratory

All samples must be collected according to professional best practices and quality control standards. The sampling company must have in place a quality objectives protocol and should submit soil samples for laboratory testing in laboratory approved sampling containers.

Selection of a laboratory in good standing with a recognised accreditation body such as the Canadian Association for Laboratory Accreditation (CALA) or Standards Council of Canada (SCC) is strongly recommended. As part of the laboratory's internal control standards their quality objectives program should include internal duplicate sample analysis (reported as relative percent difference [RPD]), laboratory control samples (LCS), matrix spikes (MS) and method blanks (MB).

Established control limits are an important step towards understanding when samples are considered accurate for use. Specifically, any RPD, LCS, MS or MB outside of control limits should be discussed for the potential effect it has on the interpretation of results.

Finally, as part of the field program, it is suggested that one duplicate be submitted for every 10 sample submissions.

6.0 STATISTICAL EVALUATION

A statistical evaluation in the context of assessing the presence or absence of statistically significant anthropogenic inputs is discussed below. The first step in a statistical evaluation is to transform the question (*i.e.*, extent of anthropogenic inputs above background concentrations) into statistical terminology, known as a hypothesis test. An overview of hypothesis testing is provided in Section 6.1. The statistical assumptions and methods applicable to determining the extent of anthropogenic inputs above background concentrations are discussed in the following subsections (Sections 6.2 to 6.7), including:

1. **Independent and Random Sampling** - data is independent when there is no correlation between data points but when correlation between the data points exists than the effectiveness of the statistical tests can be decreased. Random sampling occurs when all possible samples have the same probability of being selected.
2. **Type I and II Errors** – two types of errors are possible when deciding to accept or reject the null hypothesis. A Type I error rate is the rejection of a true null hypothesis (also known as a "false positive"). A Type II error rate is failing to reject a null hypothesis (also known as a "false negative").
3. **Sample Size** - statistical methods assume sample data are drawn from a larger population of data. Statistical methods infer characteristics of the larger population from the sample data

within a specified level of accuracy. A minimum sample size is needed to make such inferences with the specified level of accuracy.

4. **Preliminary Data Analysis: Descriptive Summary Statistics, Normal Distribution, Outliers, and Non-Detect Chemistry** - preliminary data analysis involves conducting a preliminary descriptive data comparison, verifying assumptions and identifying anomalies. Descriptive summary statistics help understand general characteristics, verify assumptions, and identify anomalies of the data set before conducting more formal statistical evaluations. Normal distribution is a bell-shaped density curve with a single peak that is symmetric about the mean and lacks skewness. Outliers are values within the data set that are statistically dissimilar from the other values. Non-detect results does not necessarily represent the absence of a chemical in a sample but infers the chemical concentration is somewhere between zero (0) and the reportable detection level (RDL).
5. **Common Statistical Errors** - various common misapplications of statistical methods to environmental data sets can occur during statistical evaluation. Failing to consider the impact of non-detect chemistry substitution methods, assuming powerful statistical inferences can be made based on small sample sizes, and the misinterpretation that a statistically significant result is equivalent to an environmentally relevant one.

For a more detailed discussion of statistical evaluations applicable to determining the extent of anthropogenic inputs above background concentrations, the ITRC's (2013) and US EPA's (2002) guidance documents include useful information.

6.1 Hypothesis Testing – Is My Site Background?

A statistical hypothesis refers to the quantitative method used to determine whether a specific statement can be supported or rejected by examining relevant data; it is evaluated through hypothesis testing. The hypothesis test will evaluate a null hypothesis (H_0). The alternative hypothesis (H_A) is the logical opposite of the null hypothesis and is accepted in the event that the null hypothesis is rejected by the statistical evaluation (US EPA, 2002).

A null hypothesis can also be termed "one-tailed" or "two-tailed" based on the number of possibilities the alternative hypothesis presents. When comparing an investigation site to background concentrations, one-tailed tests are most common (US EPA, 2002). This is simply due to the fact that we are mostly concerned with the consequence of adding (or contaminating) a Site due to an anthropogenic input. Therefore, the Investigation Site is expected to by "greater" than the background dataset, and our hypothesis is "one-tailed". If we were unsure of the data (concentrations could be "less" than or "greater" than background) the hypothesis would be two-tailed.

In hypothesis testing, two types of errors are possible (*i.e.,* Type I and Type II errors) when making the decision to accept or reject the null hypothesis. The relationship between possible decisions and errors is summarized in Table 1.

Table 1 Hypothesis Testing Error Rates

	Hypothesis	
Decision	***H_0 is true***	***H_0 is false***
H_0 is not rejected	Correct Decision	Type II Error / False Negative (β)
H_0 is rejected	Type I Error / False Positive (α)	Correct Decision

A Type I error rate is the rejection of a true null hypothesis (a false positive). The probability of making a Type I error is numerically expressed by alpha (α). Increasing alpha (α) thereby increases the probability of making a Type I error. A Type II error rate is failing to reject a null hypothesis that is indeed false (a false negative) (ITRC, 2013; US EPA, 2002).

Due to the relationship between Type I and Type II errors (as shown in Table 1), a decrease in one error rate will generally cause an increase in the other error rate. It is recommended that the selection of acceptable error rates for a Study Area hypothesis test be made on a site-specific basis.

Statistical confidence and power are known as the complements to Type I and Type II errors rates. It is also important to note that each type of hypothesis test is based on a set of assumptions (*e.g.,* normality) and the data sets to be compared must meet the assumptions of the test to be used. Failure to check the assumptions required to select appropriate statistics can result in statistical errors (ITRC, 2013; US EPA, 2002).

6.2 Independent and Random Sampling

A statistical assumption is that the data used in a standard statistical test is independent and randomly sampled. Independent and random samples are used to reduce bias and must not be positively correlated over space or time (US EPA, 2002). Data is independent when there is no correlation between data points, when correlation between the data points exists the effectiveness of the statistical tests are decreased. Random sampling occurs when all possible samples have the same probability of being selected (Frerichs, 2008).

Samples are not always independent and randomly collected. Soil data are susceptible to correlation as the samples can be spatially correlated (sampled in the same area) and temporally correlated (collected at the same time) (US EPA, 2002). Random sampling at different locations or times might not be necessary for samples with a high degree of natural mixing and homogeneity (ITRC, 2013).

Independent samples or samples that have been adjusted for correlation will fit into a standard statistical test and will ensure that samples are representative and unbiased (ITRC, 2013).

Statistically, there are few ways in which to evaluate that the dataset is truly independent and random. Therefore, the process for evaluating true independence of observations is one of assessing the quality of the data-set collected. For example, the question should be asked whether samples were collected from one location and whether that location is considered sufficiently randomly assessed. The reader is directed to review sampling design methodology in insuring independent and random sampling procedures. The US EPA (2002) document "*Guidance on Choosing a Sampling Design for Environmental Data Collection*" provides useful techniques in assuring independent and random sampling is achieved.

6.3 Sample Size

The overall goal of finding an ideal sample size for a statistical evaluation is based on the fundamental premise that statistical methods assume sample data are drawn from a larger population of data. That is, using sample data, statistical methods infer characteristics of the larger population (*e.g.,* mean concentration) from the sample data within a specified level of accuracy. A minimum sample size is needed to make such inferences with the specified level of accuracy (ITRC, 2013).

A variety of factors, such as desired statistical confidence and power, influence the ideal sample size and the relationship between sample size and each influencing parameter may vary depending on the statistical test. Sufficient sample size can also vary by statistical method.

It should be recognized that calculations involving these factors can quickly become complicated and statistical software packages can be useful for simulating sample size characteristics (ITRC, 2013). As a general go-by, sample populations inclusive of at least 10 data-points is recommended.

6.4 Preliminary Data Analysis

Before proceeding to formal statistical tests to compare an Investigation Site to background data, a preliminary data analysis of the Investigation Site and background data should be conducted. A preliminary data analysis involves conducting a preliminary descriptive data comparison, verifying assumptions (*i.e,* normality), and identifying anomalies (*i.e.,* outliers). This section provides information that is useful for parametric and non-parametric methods.

Parametric tests are based on assumptions about the parameters of a known probability distribution (*e.g.,* normal or lognormal) that from which the data are selected from. Whereas non-parametric tests do not rely on any knowledge of the type of distribution (ITRC, 2013; US EPA, 2002).

6.4.1 Descriptive Summary Statistics

The purpose of a preliminary descriptive summary statistic comparison between the investigation and background data sets is to determine important features of the data sets, including the central tendency, dispersion, and shape. Summary statistics help understand general characteristics, verify assumptions, and identify anomalies of the data set before conducting more formal statistical evaluations. It should also be noted that summary statistics are not for the purposes of formally evaluating the soil chemistry of investigation sites *versus* background.

The central tendency (*i.e.*, the central value of the distribution) can be evaluated by using summary statistics such as the mean, median and mode. The mean of a data set is the average value, the median of a data set is the middle value (when the data set is arranged in numeric order), and the mode is the most common value within the data set (ITRC, 2013; US EPA, 2002).

The dispersion (*i.e.*, spread of the data) around the central tendency and can be evaluated by the interquartile range, variance, standard deviation, and coefficient of variation. The interquartile range (IQR) of a data set is the difference between the 25th and 75th percentiles (or upper and lower quartiles). The variance is a measure of how far values in the data set are spread out. The standard deviation indicates the degree to which the values in the data set differ from the average value or mean. The dispersion across several data sets can also be compared by the coefficient of variation (CV), which is a unitless ratio of the standard deviation to the mean (ITRC, 2013; US EPA, 2002).

The shape of the data can be evaluated by the skewness and kurtosis. The skewness is a measure of asymmetry of a data set. The kurtosis is a measure of the data set's 'peak' near the mean value, compared to a normal distribution. Evaluating the shape of the data can aid in evaluating the data set for a normal distribution/normality (ITRC, 2013).

Simple graphical displays (*e.g.*, histograms) can also help visualize the central tendency, dispersion, and shape of a data-set.

6.4.2 Normal Distribution

A normal distribution is a bell-shaped density curve with a single peak that is symmetric about the mean and lacks skewness. The median value will also be equal to the mean value when a data set is normally distributed (ITRC, 2003).

In general, soil chemistry data will not be normally distributed and will require a transformation (*e.g.*, logarithmic) to fit normal distribution. A simple graphical display such as a histogram can be used to visualize the shape of the data set and determine if the data set fits a normal distribution. A histogram plots the frequency (Y-axis) of data values (grouped into specified numerical range 'bins' [X-axis]), permitting a quick comparison of the skewness and symmetry of the data set (ITRC, 2003).

Evaluating a data set for the purposes of determining if the data is best modelled by a normal distribution is more formally referred to as a test for normality (ITRC, 2003; US EPA 2002).

The overall purpose of testing for normality is to determine the type of statistical evaluation (*i.e.,* parametric or non-parametric test [Section 6.5]) to be used for assessing the extent of anthropogenic inputs above background concentrations. That is, statistical evaluations are based on a set of assumptions (*e.g.,* normality). Failure to check the assumptions of a method can result in statistical errors. For example, parametric tests are based on the assumption that data are normally or lognormally distributed, whereas non-parametric tests do not require any knowledge of the type of distribution and may be used when the data sets do not appear to fit known distributions (*e.g.,* normal or lognormal) (ITRC, 2013; US EPA, 2002).

Additional details regarding tests for normality are provided in Appendix A.

6.4.3 Outliers

Outliers are values within the data set that are statistically dissimilar from the other values. Outliers, for background soil quality evaluation, are simply defined as any value that exists outside of the acceptability range of other values in a sample population.

An outlier is not necessarily incorrect or unrepresentative of the data. Outliers can be the result of many factors, including natural variance, experimental error or due to an external influence (*e.g.,* an anthropogenic source, such as a spill). Additionally, an outlier present in a background data set may indicate that a background sample is not truly indicative of background conditions (ITRC, 2013; US EPA, 2002).

An outlier should not simply be eliminated from a data set and should generally be included in statistical evaluations unless there is reasonable evidence that they are the result of an error (ITRC, 2013; US EPA, 2002).

A simple screening method for identifying possible outliers is a box and whisker plot. The extent of the 'box' is illustrated by the 25th and 75th percentiles (or IQR). The 50th percentile (median) is depicted by a line within the box. The 'whiskers' on either side of the box are defined by 1.5 the IQR. Data value points falling outside of the range depicted by the whiskers should be further evaluated using formal statistical tests (ITRC, 2003).

Statistical software packages can be useful for testing for outliers. Failure to formally identify outliers within a data set can impact the conclusions drawn from a statistical evaluation. For example, if the maximum value (*i.e.,* an unidentified outlier) within the background data set is selected as the screening value for comparing investigation site data, the value could be unrepresentative of actual background conditions. That is, if the selected screening value is an unidentified outlier, the screening value could be much greater than warranted and the probability of identifying anthropogenic inputs above true background conditions would be reduced.

It is also important to note that parametric tests are considered more sensitive to the presence of an outlier in a data set compared to non-parametric tests (US EPA, 2002).

A detailed example of testing for outliers is provided in Appendix A.

6.4.4 Non-Detect Chemistry

Soil chemistry data sets from often present chemical results that are below the laboratory reportable detection level (RDL), which is the smallest chemical concentration that can be reliably measured. A non-detect result does not necessarily represent the absence of a chemical in a sample but infers the chemical concentration is somewhere between zero (0) and the RDL (US EPA, 2002).

It is common practice that one half of the RDL be used for reporting purposes (US EPA, 1998 and 2002; Health Canada, 2010). This substitution approach assumes that the average value of non-detects could be as high as half the detection limit. Another recommended substitution approach is to use the RDL divided by the square root of two (2) (US EPA, 2002). The approaches are more favourable than assuming non-detects as zero (0) (*i.e.*, assumes undetected chemicals are indeed absent), substituting a random value between zero (0) and the RDL, or that all non-detects are equal to their RDL (*i.e.*, the largest concentration of the analyte is being assumed present but not detected) (Helsel, 2006; US EPA, 2002).

The substitution method allows for the use of parametric methods, as parametric methods (Section 6.5.1) require a numeric value for all data points. However, substituting a false numeric value in quantitative assessment can also impact the statistical evaluation, such as affecting the estimated distribution parameters, impact a test for normality, or increasing the probability of committing a Type I or II error (Helsel, 2006; ITRC, 2013; US EPA, 2002).

If a considerable number of non-detects are present in a data set, using a non-parametric test may be necessary. Non-parametric methods can evaluate data sets with non-detects without incorporation of substitution, as non-parametric methods are based on ranking the data (ITRC, 2013; US EPA, 2002). Thus, a decision should be made on whether inclusion of non-detect data, should or should not include the substitution method.

6.5 Confidence Intervals

Confidence intervals help to quantify uncertainty about data and are a range of values that are likely to contain the statistic of interest. Confidence intervals show the upper and lower confidence limit where there is confidence that the statistic of interest will occur, such as mean, with a specific confidence level (ITRC, 2013). 95% is a common confidence interval (ITRC, 2013; US EPA, 2002). Confidence limits can be used to identify the upper and lower limits of the background values with a probabilistic statement about the mean background values (ITRC, 2013). The upper confidence limit can be used instead of the mean in comparison to guidelines. The upper confidence limit of the background can be used for data where the background mean exceeds guidelines.

6.6 Comparing Investigation Site Soil Chemistry to Background

As discussed above, parametric and non-parametric methods are the two main types of tests that can be used for statistical comparison of investigation site data to regional background data to determine the extent of anthropogenic inputs above background concentrations.

Parametric tests are based on assumptions about the parameters of a known probability distribution (*e.g.,* normal or lognormal) from which the data are selected from, whereas non-parametric tests do not rely on any knowledge of the type of distribution (ITRC, 2013; US EPA, 2002).

6.6.1 Parametric Methods

In the context of demonstrating the presence or absence of anthropogenic inputs above background concentrations, the parametric tests involve the comparison of the distribution of concentrations between the investigation site and background.

A method commonly used in parametric data analysis is the Student t-test, which is a statistical test for a difference in means (*i.e.,* difference in distribution means between the investigation site and background). A Student t-test assumes normality and equal variances. As discussed in Section 6.1, two forms of a one-tailed null hypothesis can be applied, H_0: $\mu_S \leq \mu_B$ or H_0: $\mu_S \geq \mu_B + S$. In brief, a t-statistic is calculated from the data sets, which is compared to a critical value specific to the level of significance (alpha, α). The results of the test can be summarized by comparing the p-value, which is an indication of the strength of the evidence against the pre-determined null hypothesis. That is, if the p-value is less than alpha (α), the null hypothesis is rejected and the alternative hypothesis is accepted (ITRC, 2013; US EPA, 2002). It should also be noted that the calculation of an Upper Confidence Limit (UCL) or Background Threshold Value (BTV) for direct comparison on a sample-by-sample basis can yield useful results.

If the data sets contain outliers or non-detect values, a non-parametric method may be more powerful than a parametric method for comparing the soil chemistry of the investigation site and background.

6.6.2 Non-parametric Methods

Non-parametric methods are useful when the underlying assumptions of parametric methods about the data distribution are not met. A method that is commonly used in non-parametric data analysis to determine if a difference exists between an investigation site and background is the Wilcoxon Rank Sum Test, which is a statistical test for evaluating whether data values in one data set (*i.e.,* investigation site) are consistently larger or smaller than the other data set (*i.e.,* background) by comparing the relative data ranks. This test can be used to compare two data sets that are not normally distributed and cannot be normalized by a transformation (*e.g.,* lognormal). The Wilcoxon Rank Sum Test assumes equal variance, data sets are composed of independent, random samples, and the data values within each data set are also independent of the other data values.

The Gehan's test is another non-parametric method that is recommended if there are a large number of non-detect samples (or if there are a larger number of different RDLs within a data set). The Gehan's test is considered to be a form of the Wilcoxon Rank Sum Test that uses a modified ranking procedure (US EPA, 2002).

Finally, as previously mentioned, the calculation of an Upper Confidence Limit (UCL) or Background Threshold Value (BTV) for direct comparison of Site analytical data on a sample-by-sample basis to background upper limit can also provide meaningful results. Additional details regarding parametric and non-parametric methods can be found in the ITRC's (2013) and US EPA's (2002) guidance documents.

6.7 Common Statistical Errors

The purpose of this section is to highlight the various common misapplications of statistical methods to environmental data sets that occur during statistical evaluations (ITRC, 2013; US EPA, 2002):

1. Assuming powerful statistical inferences can be made based on small sample sizes. Proper planning and establishing the number of samples to collect prior to sampling can help avoid small sample sizes and establish tolerances for Type I and II errors.
2. Failing to reject the null hypothesis automatically "proves" the null hypothesis without considering the statistical power of the test.
3. Support a test for normality with graphical displays. A test for normality is impacted by sample size and the test may not be appropriate based on the data set. Graphical displays (*e.g.,* histogram) can help support or refute the results from a normality test.
4. An outlier should not simply be eliminated from a data set as it is not necessarily incorrect or unrepresentative. Outliers should generally be included in statistical evaluations unless

there is reasonable evidence that they are the result of an error. This applies to both investigation site and background data sets.

5. Consider the impact of non-detect chemistry substitution methods as this method can affect the estimated distribution parameters, impact a test for normality, or increasing the probability of committing a Type I or II error. A non-parametric test may be required.
6. Understand the statistical parameters before conducting comparisons. Comparing unlike statistical parameters will lead to errors.
7. Check the specific underlying assumptions required for statistical tests. Conducting a preliminary data analysis can help evaluate distributions qualitatively and check assumptions.
8. A statistically significant result is not equivalent to an environmentally relevant result and the distinction between the two concepts should be addressed when evaluating the presence or absence of anthropogenic inputs above background concentrations.

7.0 REQUIREMENTS FOR REGIONAL SOIL QUALITY PROGRAM – CHECKLIST

Completion of a regional soil quality program should contain, at a minimum, all of the following information:

Scoping

1. a description of the all existing data presently available for the Site investigation;
2. a description of why a regional background soil quality assessment is required and the purpose for the assessment (*i.e.,* identification of Contaminants of Potential Concern);

Defined Study Area

3. a map illustrating the regional Study Area including: topography, surficial geology, soil types, vegetation, major land uses and direction of groundwater flow;
4. a map illustrating the area(s) where anthropogenic input may have occurred (*e.g.*, roads, historical release areas, infrastructure *etc.*);
5. a description of the parent material including soil texture, depth to groundwater and climate;
6. a description of climate, depth to groundwater and presence or absence of groundwater discharge zones, and any surface water features;
7. Slope position, or more specifically the presence of groundwater discharge zones;
8. Surface water features, particularly sloughs;
9. a description of anthropogenic activities at the Site. Specifically, a description of reported historical releases, including their location and the status of any subsequent environmental assessment or remediation;

Data Quality Assessment

10. a brief discussion of the quality assurance and quality control method(s);
11. calculation of required sample size;

Statistical Evaluation

12. results of a preliminary data analysis (test for normality, check for outliers);
13. decision on how to compare Study Area results with investigation site data;

Results and Discussion

14. tables presenting all analytical data;
15. a discussion of the result of the statistical evaluation, interpretation of background chemistry (*e.g.,* range, upper control limit, *etc.*);
16. a discussion of whether the background soil quality is suitable for direct comparison to investigation site(s); and
17. clear indication of whether investigation site data is considered to be within background.

8.0 ALTERNATIVES TO THE REGIONAL METHODOLOGY

In cases where the Site conditions are not consistent with the regional conditions, or the Site does not fit into an existing Domains, the Site will generally be assessed individually using Site specific background data and input parameters. Other alternative assessment methods may also be considered on a Site-by-Site basis.

9.0 CONCLUSIONS

The methodology presented in this report was developed for assessment of salinity impacted sites in southern-Alberta. However, the example method is also appropriate for use with any parameter where background concentrations have been shown to influence policy decision (*i.e.,* parameter concentrations may naturally occur at levels above guideline criteria).

The process presented is broken down into four (4) steps: Scoping, Defined Study Area, Data Quality Assessment and Statistical Evaluation. Together, the purpose of these steps is to assure that samples are collected from meaningful locations and that the samples are free of erroneous data. That data is then pooled (where statistical appropriate to do-so) and builds a Regional Background Data-Set. Each data-set becomes a resource by which Investigation Sites meeting the conditions of similar morphology and soil type can be routinely compared. When the regional background data set is applied to one or more investigation sites the proposed methodology will reduce the data collection requirements for the region (having a centralized background Study Area) and allow for more efficient remediation of contamination, reducing both economic costs and environmental impacts.

10.0 REFERENCES

Alberta Agriculture and Forestry. 2004. *Salinity Classification, Mapping and Management in Alberta.*

Alberta Environment and Parks (AEP). 2016a. *Tier 1 Soil and Groundwater Remediation Guidelines*. February 2016.

Alberta Environment and Parks (AEP). 2016b. *Tier 2 Soil and Groundwater Remediation Guidelines*. February 2016.

Alberta Environment (AENV). 2001. *Salt Contamination Assessment & Remediation Guidelines*.

Ayres, K.W., Acton, D.F., and J.G. Ellis. 1985. *The soils of the Swift Current Map Area 72J Saskatchewan*. Extension Division, University of Saskatchewan, Saskatoon.

Canada – Alberta Environmentally Sustainable Agriculture Agreement (CAESA). 1998. *Soil Inventory Project Procedures Manual.*

Chang, C., G.C. Kozub and D.C. MacKay. 1985. *Soil salinity status and its relation to some of the soil and land properties for three irrigation districts in southern Alberta*. Canadian Journal of Soil Science 65(1): 187-193.

Canadian Council of Ministers of the Environment (CCME). 2016. *Guidance Manual for Environmental Site Characterization in Support of Environmental and Human Health Risk Assessment, Volume 1 Guidance Manual.*

Delin, G.N. and D.W. Risser. 2007. *Ground-Water Recharge in Humid Areas of the United States--A Summary of Ground-Water Resources Program Studies, 2003-2006*. Groundwater Resources Program. Unites States Geological Survey.

Dyck, M.F. 2001. *Long-term solute transport under transient, semi-arid conditions*. M.Sc., Thesis Report, University of Saskatchewan. 202 pp.

Dyck, M.F., Kachanoski, R.G., and E. de Jong. 2003. *Long-term movement of a chloride tracer under transient, semi-arid conditions.* Soil Sci. Soc. Am. J. 67: 471-477.

Ecological Stratification Working Group. 1995. *A National Ecological Framework for Canada.* Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, Ottawa/Hull. Report and national map at 1:7500 000 scale.

Environment and Sustainable Resource Development (ESRD). 2014. *Subsoil Salinity Tool.* Software development by Equilibrium Environmental. Developed for Alberta Governemnt.

Frerichs, R.R. 2008. Simple Random Sampling. Rapid Surveys 3: 1-44.

Gee, G.W., J. Keller and A. Ward. 2005. *Measurement and Prediction of Deep Drainage from Bare Sediments at a Semiarid Site*. Vadose Zone Journal 4(1):32-40

Gee., G.W. and D. Hillel. 1988. *Groundwater recharge in arid regions: Review and critique of estimation methods.* Hydrological Procedures. 2:255-266.

Government of Alberta, Land Use Section. 2016. *Canada - Alberta Environmentally Sustainable Agriculture Agreement (CAESA): Soil Inventory Project Procedures Manual* - Data Dictionary. Www1.agric.gov.ab.ca. N.p., 2016. Web. 18 Feb. 2016

Helsel, D. 2006. Fabricating data: How Substituting values for nondetects can ruin results, and what can be done about it. Chemosphere 65 (2006) 2434–2439.

ITRC (Interstate Technology and Regulatory Council). 2013. *Groundwater Statistics and Monitoring Compliance: Statistical Tools for the Project Life Cycle.* Groundwater Statistics and Monitoring Compliance Team, Washington DC. December 2013.

Natural Regions Committee. 2006. Natural Regions and Subregions of Alberta. Compiled by D.J. Downing and W.W. Pettapiece. Government of Alberta. Pub. No. T/852.

Sample, B.E., Lowe, J., Seely, P., Markin, M., McCarthy, C., Hansen, J., and Aly, A. 2014. *Depth of the Biologically Active Zone in Upland Habitats at the Hanford Site, Washington: Implications for Remediation and Ecological Risk Management*. Integrated Environmental Assessment and Management. 11:150-160.

Schreiner, B., 1990: *Lithostratigraphic correlation of Saskatchewan tills – A mirror image of Cretaceous bedrock.* Saskatchewan Research Council Publication No. R-1210-3-E-90, report volume 1.

bedrock. Saskatchewan Research Council Publication No. R-1210-3-E-90, report volume 1.

US Environmental Protection Agency (US EPA). 2002. *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*. Office of Emergency and Remedial Response, Washington DC. September 2002.

Vaccaro, J.J. 1992. *Sensitivity of groundwater recharge estimates to climate variability and change, Columbia Plateu, Washington*. Journal of Geophysical Research Atmospheres 97(D3):2821-2833

Wiebe B.H., Eilers R.G., Eilers W.D., and J.A. Brierley. 2005. *Soil Salinity: Risk of Soil Salinization.* Environmental Sustainability of Canadian Agriculture: Agri-Environmental Indicator Report Series—Report #2. Chapter 15. 114-118.

Wilson, M.J. 2006. *Factors of soil formation: Parent material. As exemplified by a comparison of granitic and basaltic soils*. P. 113-129.

Woods, S.A., R.G. Kachanoski, and M.F. Dyck. 2006. *Long-term solute transport under semi-arid conditions pedon to field scale*. Vadose Zone J. 5:365-376.