APPLICATION NOTE

How many soil water content measurements are enough?

CAMPBELL SCIENTIFIC, INC.

How many soil water content measurements are enough?

This application note provides basic statistical methods used to determine the number of measurements needed to provide a specific level of confidence that the soil water content of a given area has been adequately characterized.

Spatial Variability of Soil Water Content

Soil water content can vary significantly among several locations which are near each other and apparently similar. Water content measurements using the most accurate methods available provide evidence of significant differences in soil structure and texture even when the measurements are limited to an area of only one square meter. The degree of variability is dependent on many factors including presence of coarse fragments (rocks), micro and macro-faunal activity, erosion, management practice, and plant root activity. The difference in soil physical properties from location to location and the subsequent difference in soil water characteristics is commonly referred to as spatial variability of soil hydraulic properties. Any characterization of an area used to manage or inventory water resources in that area must account for this variability.

Example Data Set

As an example, Table 1 lists the results of 12 volumetric water content measurements taken within a two meter radius on a well-established and level lawn. Below the table are the mean, standard deviation, and range of the table's data.

Table 1. Sample Water Content Data

10.7	7.4	12.7	12.5
11.9	12.5	12.1	14.0
13.4	12.5	10.2	12.5

mean   11.9
stdev   1.74
range   6.6

Confidence Intervals

Statistical approaches are based on probability theory. Using confidence intervals is a method of expressing probability. Consider measuring the water content at several different locations in a volume of soil. The measurements will differ in value but will be distributed about a mean value. If another measurement is then taken, you can calculate the probability that this measurement will fall within a specific amount below and above the mean value. Conversely, for a specific probability value, you can calculate the water content range that the measured value will likely fall within. The sum of the values that lie below and above the mean value is known as the confidence interval. The confidence interval becomes smaller as the probability increases.

The relationship between confidence interval and probability can be described using the following expressions.

$$\mu - k \; \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + k \; \frac{\sigma}{\sqrt{n}} \qquad [1]$$

Where μ is the actual population mean, σ is the actual population standard deviation and n is the number of values used to calculate the mean, $\bar{x}$. The confidence coefficient, k, is used to specify a probability value. A probability of 95% or 0.95 could be specified and [1] would then define the range below and above the mean that is the confidence interval. The probability is 95% that a measured mean will fall within the defined interval.

Equation 2 is algebraically identical to Equation 1, but the rearranged form now describes the likelihood that the interval will bracket the actual population mean. The difference is subtle but important.

$$\bar{x} - k \; \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + k \; \frac{\sigma}{\sqrt{n}} \qquad [2]$$

Equation 2 is based on the assumption the actual population mean and standard deviation are known. Actual values are seldom known. If the values were known, you wouldn't have to measure them. Assuming a normal distribution, correcting for the fact that the actual population standard deviation, σ, must be estimated from measurements, and applying the theory of the Student's t-distribution, Equation 2 becomes:

$$\overline{x} - t\,\frac{s}{\sqrt{n}} < \mu < \overline{x} + t\,\frac{s}{\sqrt{n}} \qquad [3]$$

Where s is the estimate of σ that's based on the measurements and t is the value obtained from a table of Student's t -values. Using the Student's t-distribution corrects the normal distribution for the number of measurements taken or degrees of freedom.

Consider the data presented in Table 1. If a probability value of 0.95 is acceptable, Equation 3 can be used to define the confidence interval. For this data, the mean is 11.9 and the standard deviation is 1.74. With 11 degrees of freedom, the Student's t value is 2.201. The Student's t-distribution is symmetrical, and the probability of 0.95 implies 0.025 is on both sides of the mean. For most tables, to choose a t value for 0.95, you use the value listed for 0.975 (i.e., 1.0-0.025). However, some tables do it differently and you'd use the value listed for 0.025. Entering those values into Equation 3, you get confidence interval values of 10.79 and 13.00. In summary, there is a 95% probability that the mean water content lies between the water content range of 10.79% to 13.00%. For this example, the depth of water associated with the range is 0.44 cm in a 20 cm profile.

Minimum Number of Samples

When the the sample mean and standard deviation are known and a confidence interval is chosen, Equation 4 can be used to determine the minimum number of randomly taken measurements.

$$N = \frac{4t^2s^2}{L^2} \qquad [4]$$

Where N is the number of water content measurements, and L is the acceptable water content range as defined by the user. For example, the acceptable water content range might be $\pm 2.5\%$ so L would be 5.0.

Equation 4 relies on a good estimate of the population standard deviation. A useful approach of estimating this is to take several water content measurements then calculate the standard deviation for these measurements. This standard deviation estimate and the associated t value are used in Equation 4. The range is specified by the user and the calculated N value defines the required number of measurements. Using the same s and t values, the number

of measurements can be determined for several water content ranges.

$$N = \frac{4t^2s^2}{L^2} = \frac{4 \cdot (2.201)^2 \cdot (1.74)^2}{(5.0)^2} = 2.35$$

The calculation of N using the sample standard deviation shows that three measurements are required to give a 95% probability that the sample mean gives the desired ±2.5%.