Confidence Limits in Temperature Measurements

By
G. Raymond Peacock
Temperatures.com, Inc.
Southampton, PA 18966-3836

Abstract

The R&R (Repeatability and Reproducibility) of temperature-measuring devices applies to quantitative thermal imagers as well as the most precise temperature sensors used in standards calibration laboratories. Once you understand what's involved with R&R and how it can affect the results of your measurements, you will think about real temperature and temperature difference measurements in a new way. The links to calibration and traceability are then relatively easy steps to take. The significance of calibration of a temperature-measuring thermal imager and the likely uncertainty of results in the field begin to make real sense. A better understanding of these measurement fundamentals can help you relate measurement results and their confidence limits.

Introduction

Measurement science, called metrology by some, is a very precise discipline. It is best known for its use in National Standards labs, like NIST (National Institute of Standards and Technology) in the USA, and NRC (National Research Council) in Canada. When using measurement science, people are usually pushing the limits of their available technology to get the smallest measurement uncertainties possible. However, just because Thermographers are not, or don't think they are, pushing the limits of available technology when measuring temperatures, it does not imply that they should be neglecting good measurement science practices in their work. Measurements are measurements regardless of who makes them and they have value depending upon the understanding and necessary care taken when the measurements are made. If you report measurements, you are in the measurements business and you should understand not only your equipment, and all the lore of thermography, but also about measurement science and the use of statistics. Actually, with software and compact computers available today, the statistics are the easy part. The hard part is deciding to follow the established practices related to good measurement science practices.

The object of this paper is to review some simple measurement science concepts and how they can be used in making thermographic temperature measurements and what needs to be reported of the data taken and the people and equipment making the measurements. An instrument reading of temperature needs to be well understood and sometimes challenged by the person responsible for the measurement or else the value and confidence in the measured values are greatly diminished. Confidence is, after all, one of the keys to customer satisfaction. If they are confident that you are doing your job correctly then your relationship will grow. Similarly, confidence in measurement results is a key to self-assurance; further, it can be quantified, or not, as part of the measurement practices followed.

It is also critical to realize that better measurement practices are an integral part of ISO 9000 and all modern statistical process control and maintenance reliability practices. The quality assurance wheel is still turning, even though it doesn't make much press. Its impact will increase rather than decrease in the future, if, for no other reason, than increasing global competitiveness.

Basic Measurement Concepts

Measurement devices, such as rulers, thermometers, pressure gauges and quantitative thermographic imagers, perform measurements. All measurements contain some error; one of our important tasks is to understand and determine the size of the error and how it can be reported as part of measurement results.

Measurement results can never be better than the basic measurement capability of a given measurement device. It is often overstated, by implication, in reported results by having too many significant figures in the results data. If a result is an average of say six measurements that mathematically work out to 23.33 °C and that precise value is reported, it implies that we have a measurement capability of 0.03 °C! We may be able to see a 0.1 °C temperature value, but certainly not 0.03 °C! So, common sense when reporting result values is important and should not imply that you have more capability than is true.

As an example, typical rulers used in carpentry are graded in $1/16^{th}$ inch intervals. If one claimed a measurement capability of $1/64^{th}$ of an inch with such a device, far better than its minimum measurement resolution, it would of course not be believed because it is impossible to achieve. Furthermore, anything less than $1/16^{th}$ is suspect because that value appears to be the basic calibration limit of the device. We don't usually have rulers certified and calibrated at the $1/16^{th}$ inch level, but there are indeed gauge blocks and precision gauges used by machinists that are not only certified, but carry a correction in a certificate as a function of the block's temperature, to correct for any expansion or contraction. Typically such blocks and gauges measure to within $1/10,000^{th}$ of an inch or thereabouts. However, someone who used them would not claim measurement capability to within $1/100,000^{th}$ of an inch.

What about thermal imager temperature resolution? We can see usually 1 °C or, on some units, 0.1 °C resolution. Does that imply a measurement capability? Some manufacturers, by implication, suggest that you can, when in fact you cannot. Most thermal imagers have, as a minimum, about a 2% accuracy specification, or something closer to about 2 or 3 °C calibration uncertainty. Clearly, such devices are different, as measurement devices, than common rulers. They have a calibration limit that is larger than the temperature resolution capability of the device. So, what would be the minimum believable temperature resolution? This depends on a few other things, but certainly no better than the manufacturer's calibration specification. We'll get back to those later. Typically the result of careful measurements is reported as a number, plus or minus an uncertainty value or a standard deviation value for the data set used to calculate the estimate. Say, for discussion's sake, that the average measured value is 87.677777 °C ± 1.833 °C, where the 1.833 °C is further specified as the estimated standard deviation. The technically correct way to report these values for an instrument having a fundamental measurement capability of ±2 °C would be to round the values up to the nearest increment of resolution capability, or as: 88 °C ± 2 °C.

Since a thermal imager is an expensive, complex temperature measurement device, it is a primary, essential requirement that the measurement calibration uncertainty is well known, traceable, and its measurement stability known usually by a calibration history record. An expensive instrument without regular, periodic checks of one of its key capabilities is a wasted resource. If it is a prime source of income or plant evaluation, then you need to be sure that it functions at its best at all times.

Many Thermographers will fall prey to the argument that is often made that they are not really measuring temperatures; they are measuring temperature differences in a scene. Therefore

absolute calibration of a thermal imager is not a problem or concern. To any paranoid ear, that sounds like an excuse for not understanding how an instrument functions. The fact is, like so many "sales pitches", there is an appeal to the argument, but it is seldom true.

There are two very important aspects of instrument performance that bear on the subject of calibration stability and uncertainty whether measuring temperatures or temperature differences:

> 1. The error in a measured temperature level varies with both errors in the instrument zero and gain values, whereas errors in temperature differences vary with the error in the calibration gain and not the zero level. If the calibration gain is off, then there will be a temperature level sensitivity in measurements of true temperatures and temperature differences or gradients. For instance, say the temperature difference in a scene between two points is 20 °C. One part is at 120 °C, the other at 140 °C. Now suppose that the system zero calibration has shifted by 30 °C. In that case the difference is still 20 °C. (Most people associate zero shifts that with a temperature difference-but, in fact gain shifts are just as likely to occur and are the source of serious measurement errors.) If the system gain has shifted, the difference will vary according to the amount of the gain shift. Take the same example where the output is related by a typical linear relationship; for example, where we assume that the bias is 0.0 and the gain is 10.0:

$$\text{Output} = \text{gain} \times \text{Input} + \text{bias}$$

Factors: (bias=0 gain = 10)	Inputs before	Output Before	Output Difference	Input After	Output After	Output Difference
Change Zero by 10%	12, 14	120 °C, 140 °C	20 °C	12, 14	130 °C, 150 °C	20 °C
Change Gain by +10%	12,14	120 °C, 140 °C	20°C	12,14	132 °C, 154 °C	22 °C

Table 1-Output effects of zero and gain changes.

> If the gain shifts by +10%, a 20 °C difference will look like a 22 °C difference. It's actually a lot worse than the example given, because thermal imager calibration is not linear, it is noticeably non-linear and gain calibration errors result in much larger temperature errors.

> 2. Knowing that your calibration is good under the fixed, stable conditions of a calibration environment is not enough. If you measure the same object in an hour or a day or a month from now, chances are very good that the measurement conditions, and possibly even the person making the measurements will not be the same. You need to know the calibration stability and the effect of each of the variables that can influence the measurement results. You need to be aware of how a measuring instrument behaves when conditions that could influence its measurements change.

One set of "simple" tests for stability and calibration checking is given for spot radiation thermometers in ASTM Standard E1256. It's a good starting point for testing the stability of thermal imagers although more complete practices need to be available. Work on them has begun in ASTM Subcommittee E20.02, Radiation Thermometry. So, in order that a temperature differential measurement at one point in time can be compared fairly to another requires that the instrument be calibrated during both sets of measurements and that the effects of the likely influencing factors, that may be different each time, be known and any corrections carefully made.

Making absolute temperature measurements is yet another step in complexity, but it has the very same basics as a differential temperature measurement. Knowing that your calibration is correct

is but the first step. You need also to know the effect of the major influencing factors involved in making practical measurements. You learn about and understand measurement science. It stands to reason that if you are reporting measurements that you understand their believability. Also, your calibration checking procedures need to have a root source that is at least 4 times better than the calibration sensitivi ty you are seeking.

As an example, consider the case where one uses the boiling point of water as a "reality check" on one equipment calibration. Unless one uses traceably calibrated thermometer to verify the boiling temperature of water, one must be careful to correct for local air pressure since the boiling point is pressure sensitive. Normal weather-related atmospheric air pressure variations introduce about a 0.8 °C uncertainty in the boiling point and the altitude at which the water is boiling can introduce an even larger error. The boiling point of water changes about –1 °C for each 355 meters increase in altitude. In fact, a boiling point apparatus makes a pretty good altitude meter. You need to know how your instrument calibration is established and maintained. If you use boiling water without an independent, reference temperature sensor to indicate the actual boiling point, you can expect that your system calibration to have an uncertainty of at least 3 °C, assuming you correct for altitude effects. It's more if you don't!

Now, assuming that you have a calibrated instrument and go into the field and make a temperature difference measurement, is one measurement enough? How many is enough? Do you know the ambient temperature, the atmospheric humidity level, the solar intensity, the temperature of the objects surrounding your measurement spot, who is operating the unit, what the various instrument settings are? Good, glad you do. Do you also know the impact each of these factors has on your resulting measurement values? Unless the manufacturer of the equipment provides that relationship, you will need to test the equipment yourself, or have it tested by a qualified laboratory. We recommend that you use some established practices as, for example, recommended in ASTM E 1256.

How well do you know the thermal settling time of your imager, say, when leaving an air-conditioned vehicle and walking into an area at an ambient that is 30 °F hotter? How does your thermal imager correct for the fact that it stabilizes (in one or two hours more or less) at a temperature that is 30 °F hotter than the temperature at which its calibration is certified? Do you know? If you don't, you could be making significant measurement errors.

Characterizing the performance sensitivi ty of an imager is a matter for experts, especially the equipment manufacturers. If suppliers expect you to believe that instruments have a certain measurement capability, they should be following the same basic measurement principles that you need to follow in reporting results. They know, or should know, and be able to explain to you, the measurement capability details of their equipment in numerical terms. You may have to request the information because it is usually not included as part of the equipment specifications. In fact, the specifications produced by most imager makers are often vague and incomplete, leaving much to the imagination of the user. Part of the problem with imager measurement specifications is that the devices were developed as imaging devices and not quantitative measurement devices. The only measurement specifications that are of value for understanding measurement capabilities are those that come complete with uncertainty values at stated confidence levels under stated conditions of measurement. For example, temperature calibration is often expressed as accuracy. The preferred technical term is uncertainty, not accuracy, and it should be expressed in the same terms as NIST uses in expressing measurement uncertainty. NIST's booklet, *NIST Technical Note 1297*: *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results* includes an explanation of how they use the term. The booklet can be downloaded from the Web and is free by mail also.

Basic Measurement Statistics

Measurements of objects having temperature variations made with devices that are slightly imperfect require that an average measurement be determined. Individual measurement results are, in reality, samples from the range of possible values that the instrument reports. There is a true average value and some variability about that average. If we take only one measurement, we could be anywhere within the range of possible values. However, if the factors causing the fluctuations are random, then the effect of making additional measurement is well known and explained in simple statistics. An excellent reference to both measurement statistics and temperature measurement and calibration is the book <u>Traceable Temperature</u> by J.V. Nicholas and D.R. White, (John Wiley & Sons). Some of the important definitions used in statistics are defined in Table 2 below. Please note that a major shift has occurred in US industry over the last 10 years or so. Measurement practices are being tightened up in all industries as ways to help improve global competitiveness. The techniques and resources are well established since they have been practiced without interruption by the military, power-generation and aerospace industries since the 1950's.

Item	Definition or Source
Mean or Average	$T_{av} = (T_1 + T_2 +.. + T_n)/n$
Estimated Standard Variance	$s^2 = \{(T_1- T_{av})+(T_2- T_{av}) +..(T_n - T_{av})\}^2 /(n-1)$
Standard Uncertainty	u_c = Square root of (s^2)
(Student)t-Statistic	k from a table of t-values vs. (n-1) and p
Expanded Uncertainty	$U = k * u_c$

Table 2-Measurement Terms & Statistics

Confidence Limits and Levels

The resulting confidence limits, the real object of this paper, and level of confidence are directly related as shown in Table 3. They are based on the variability in measurement results. The confidence limits are related to the size of the standard variance and uncertainty. The confidence level results from the statistics of random errors and describes the percentage of readings that will be within the desired confidence limits.

Confidence Limits (around the Mean)	Confidence Level
$\pm\ u_c$	68.3%
$\pm 2\ u_c$	95.5%
$\pm 3\ u_c$	99.7%

Table 3- Confidence Factors

So, how does one achieve the confidence levels in temperature and temperature gradient measurements with a quantitative thermal imager? It is a big question and one that cannot be answered quickly or easily because of the many factors involved. However, the steps to obtain the limits are rather straightforward and can be easily outlined. There are two big steps with lots of little details to be acquired in the first.

> Step 1: Determine the confidence level that you can achieve in measurements in the field. That involves knowing your equipment's calibration uncertainty and its likely measurement uncertainty under less than ideal conditions. We've touched on that, but

there is also a series of tests called R&R tests to measure the influence of the equipment operator(s) on the measurement results. If properly done, the field variability sensitivity and the operator influences can be grouped together in one set of tests.

Step 2: Determine the confidence level that your customer requires. If the two levels do not match at the outset, you could be in trouble or in roses, depending on which is larger.

If you are in "trouble" there are two options, they are:

Option 1: If the customer requests smaller measurement uncertainty or better capabilities than you can deliver, one could explain that your measurement capabilities are as you have measured and documented and represent a realistic appraisal of the capability of state-of-the-art equipment and trained operators.

This option, of course, assumes two things: first, that your conclusions are true, backed by documentation and second, that the customer may be seeking unrealistic measurement performance. You should be able to convince the customer that you are competent and request that similar documentation be provided from any competitor. It doesn't always work, partly because some customers refuse to become better educated, and also when the requirements are really better than your capability.

Option 2: If the customer really needs measurements better than your best capabilities, you could undertake improving them. Having assessed your present capability carefully, you would have a very good idea of where to begin such improvements and what the cost tradeoffs would be.

But there is one more step to be considered before you have a complete understanding of your measurement capabilities or confidence limits, i.e., the combined effects of instrument errors, operator skill and measurement condition influences.

Repeatability and Reproducibility

One way the overall effects of instrument calibration uncertainties and other variances due to operators can be determined is through a set of controlled R & R tests, or Repeatability and Reproducibility tests. The basics of R&R testing lie in statistical results from controlled tests. There is a well-defined formalism used, for example, by The Automotive Industry Action Group, AIAG. They are one of the biggest driving forces (no pun intended) in improving production quality in North America and have published a series of booklets and practices recommended for measurements and measurement devices. Any company expecting to do business with a major auto manufacturer or their suppliers in the USA, Canada or Mexico must follow these practices in order to be a minimally acceptable supplier.

Included in AIAG's basic measurement quality assurance are R&R measurement procedures for testing equipment and operators. Although written primarily for dimensional gauging (a significant portion of automobile production quality requirements), the practices are applicable to any measuring device. Within the automotive industry, this type of testing is often called GRR, standing for Gauge R&R. The handbooks and sample data sheets are available at modest fees from the AIAG and some of the software vendors to the industry.

Within the semiconductor industry, a very similar need was evident. They worked with their own research and production resources and NIST to develop a measurement practices policy that follows the same methodology as the AIAG's. The resulting measurement practices handbook is

freely available on the Web and can be viewed and downloaded from the NIST web site. The version of the Handbook on the NIST web pages is integrated with the Dataplot statistical software. In order to use Dataplot from the Handbook, it must be downloaded and installed on your computer.

The basic procedure for R&R testing is also straightforward. One starts with a calibrated measurement device and has an operator measure a variety of objects, usually about three to five, each having a different value, that are different but not necessarily known. The only requirement is that they do not change during the tests. Several operators using the same objects perform the same set of measurements, usually with only one instrument shared among them. That corresponds to one testing round. Then the round is repeated, usually two to five times.

If different environmental conditions are likely to affect the results, then one or more of the objects can be in different real or simulated environments. The key is to have each operator or "appraiser" repeat the same measurement more than once, usually a minimum of two or three times. Each of the operators measures the same objects. This enables one to statistically evaluate and separate the effects of the operator, the effects of the equipment and the effects of the environment. It also enables one to determine the statistics related to the combined effects of all the major influencing factors.

There are numerous software packages on the market as well as a complex, but free package, Dataplot, from NIST that will not only help one set up R&R tests, but can also guide one through the test steps and calculate the resulting statistics from the measurement data.

Conclusions

1. Temperature measurements made with thermal imagers are like any other measurement; they have built-in errors.
2. There are well-established methods for assessing such errors and reporting measurement results with confidence limits to meet the users expected measurement confidence levels.
3. It is in your best interest to begin to practice good measurement science in order to responsibly qualify your measurement capability and measurement results with confidence factors that can enable you to meet the expectations of your customers.

If you don't follow good measurement practices you will lose to the supplier that does.

References/Resources

ASTM Standard E1256*, Standard Test Methods for Radiation Thermometers (Single Waveband Type).* W. Conshohocken, PA: American Society for Testing and Materials, 1995. (On the web at http://www.astm.org - downloadable for a small fee)

Nicholas, J.V. and D.R. White, *Traceable Temperatures Second Ed.*, John Wiley & Sons, Ltd., 2001.

Taylor, B. N. and C. E. Kuyatt. *NIST Technical Note 1297:Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results,* National Institute of Standards and Technology, US Dept. of Commerce, Gaithersburg, MD, 1994. (On the web at: http://physics.nist.gov/cuu/Uncertainty/index.html).

AIAG-Automotive Industry Action Group, *MSA-3 Measurement Systems Analysis (MSA) Third Edition for Automotive QS-9000 Suppliers*, 2002, (Can be purchased by telephone from AIAG Customer Service department at (248) 358-3003 or on the Web at: http://www.aiag.org/publications/quality/iatfquality.html) Automotive Industry Action Group (AIAG), 26200 Lahser Road, Suite 200, Southfield, MI 48034

Engineered Software, Inc. *Measurement Assurance* (Software supporting the analytical techniques detailed in the Automotive Industry Action Group (AIAG) Measurement System Analysis Manual), Engineered Software, Inc., 43737 Timberview Drive, Belleville, MI 48111 (On the Web at: http://www.engineeredsoftware.com)

Croarkin, C., Editor, *Measurement Process Characterization*, Chapter 2 in the NIST/SEMATECH e-Handbook of Statistical Methods, National Institute of Standards and Technology, US Dept. of Commerce, Gaithersburg, MD (On the web at http://www.itl.nist.gov/div898/handbook/index.htm)

Filliben, J. J. and A. Heckert, *Dataplot* (A free, public domain, multi-platform {Unix, VMS, Linux, Windows 95/98/ME/XP/NT/2000, etc.} software system for scientific visualization, statistical analysis, and non-linear modeling with GUI interface by R. R. Lipman), National Institute of Standards and Technology, US Dept. of Commerce, Gaithersburg, MD 1978-2002 (On the web at http://www.itl.nist.gov/div898/software/dataplot/)