Special Issues and Methods in Testing
Uncertainty Evaluating Software

May 19, 2005

Dr. Craig Shakarji
Manufacturing Engineering Laboratory
National Institute of Standards and Technology
A Bad Day at Work?

- You come into work to find a software CD on your desk
- It claims to evaluate the (task specific) uncertainty of measurements (for example, measurements made by CMMs)
- You are tasked with testing to find out whether this software works (Can you either “certify” this or prove it to be wrong?)
- You begin to wonder what bad thing you did to your supervisor to deserve this
Some CMM Types

- Moving bridge
- Fixed bridge
- Column
- Fixed table cantilever
- Moving ram horizontal arm
- Moving table horizontal arm
- Gantry
- L-shaped bridge
First Attempt

• Measure the length of a gauge block with a CMM

• Gauge calibrated to be 100.000 mm (with small calibration uncertainty)

• Measure to find CMM reporting 100.030 mm

• Plug problem into software and find reported $u = 0.012$

• ??? Cannot say much! $u$ is a statistical value
Second Try

- Measure a calibrated artifact a statistically meaningful number of times in several positions and orientations
Second Try

- Measure a calibrated artifact a statistically meaningful number of times in several positions and orientations
- Excuse me—may I borrow 100 artifacts, 100 CMMs and 1,000,000 hours of operation?
An attempt at comprehensive coverage would involve:

- Various sized artifacts
- Various geometries (planes, cylinders, spheres, cones, complex surfaces)
- Various shapes (cylinders of various aspect ratios, cones of different apex angles, etc)
- Various positions, orientations
- Various numbers of points sampled and sampling strategies
- Various CMMs
- Etc. (Depends on what is claimed by the software)
Next Idea: Reference Values

• Somehow come up with the “right answer” to a specific uncertainty evaluation and compare the reference value directly to the software’s reported value.

• This would solve the problem of the uncertainty being a statistical value requiring a statistically meaningful number of comparisons. It would reduce to a single comparison.

• Envision:

\[ u_{\text{ref}} = 0.012 \quad u_{\text{reported}} = 0.012 \quad \text{Good!} \]

\[ u_{\text{ref}} = 0.012 \quad u_{\text{reported}} = 0.02 \quad \text{Bad!} \]
Problems

• Can obtain reference values that match reference measuring situations for few cases

• A more subtle problem is that uncertainty is based on level of knowledge—how can we know these are the same?
Knowledge / Probability Interplay

- Two bags, A & B. A man goes to one bag and selects one chip.
- Bag A contains 90 blue chips and 10 red chips
  Bag B contains 10 blue chips and 90 red chips
- What is the probability selected chip is blue?
- Only knowing this information, probability is 50%
- If you also saw that he picked from Bag A, the probability you would report is 90%
- If you also peeked and saw a part of the blue chip in his hand, you would report a probability of 100%
Knowledge / Uncertainty Interplay

- The same is true of CMM measurement uncertainty
- The UES gains knowledge of CMM through
  - A single MPE number
  - Five performance evaluation numbers
  - A query using several measurements with a ball plate
  - An extremely detailed series of measurements to understand the CMM behavior in very specific detail
- These increasing levels of knowledge can lead to lower uncertainties reported – but all correct!
  - This is not overestimating uncertainties
  - Like the probability of having drawn a blue chip, two different answers can both be correct, as they are based on different knowledge
Reference Values for Uncertainty (cont)

- In some cases, a single reference value is not well defined, since the “correct” reference value is actually a function of the knowledge gained about the measuring situation.
- The dependence on knowledge applies to more than the CMM.
- Even without this knowledge issue, measuring situations for which a **reliable** reference value is known are very simple cases.
- You now realize you have been given a very difficult task.
So far …

• Physical testing is too burdensome
• Reference values can be ill defined and may be reliably available only for simple measuring situations
• However … Both serve as good necessary conditions. That is, passing some of these tests might give assurance that the software is good, but failure in these tests can show the software problematic
Another Method: CVE

- Computer-aided Testing and Evaluation (CVE)
- Use a computer to simulate a measuring situation (including the CMM itself)
- Since the error behavior of a CMM is really a vector field over the measuring volume (dependent on time), this can be simulated by a computer.
- The information gathering required by the UES can also be simulated using the CMM mathematical definition
- The measurements can also be simulated—again using the mathematical definition of CMM
CMM Population

Select One CMM

Perform CMM Assessment

Create Measurement

Create $U$ Statement

Calculate Meas. Error

Compare: $|E| < U$?

Yes = "successful"

No = "unsuccessful"

Success Statistics

Repeat process with new length

Are 95% of Meas. Successful?

Yes = "Pass" this CMM

No = "Fail" this CMM

Pass/Fail Statistics

Repeat process with new CMM

Report % CMMs Pass & Lengths Successful
Summary of Methods

- Physical testing, reference values, and CVE
- All have advantages and disadvantages
- None comprehensive in itself
- Using a combination can increase confidence in software, even if “certification” or essentially guaranteed performance is out of reach
- Methods described in more detail with more advantages and disadvantages should be given in ISO TR 15530-4 (currently in late stages of development)
Method 1 – Physical Testing

• **Advantages**
  – This testing matches what the UES will actually be used for.
  – Testing is performed on a real-world CMM—the one of interest
  – Testing includes the gathering of the input data from the machine

• **Disadvantages**
  – Physical measurements are time consuming and costly. Several measurements are needed to establish reliable, statistical comparisons.
  – Testing a wide variety of measurands requires many calibrated artifacts. (For example, just using cones, one might need a variety of apex angles, sizes, and aspect ratios.)
  – Testing on one CMM does not guarantee that the UES will work on other CMMs, since their error sources might be different.
  – UES usually does not consider all influence factors, so it is hard to tell if errors larger than the reported uncertainty are due to other factors or a problem in the UES
Method 2 – Comparison With Specific Reference Values

- **Advantages:**
  - The comparison is directly between two uncertainty values
  - Using very restricted measuring situations (unrealistic in physical testing) can help identify specifically where problems occur in the software

- **Disadvantage**
  - The test only applies to very restricted situations covered by the software under test, because it is hard to obtain reliable reference values for complex situations
Method 3 – CVE

**Advantages:**

- A large number of simulated measurements can be carried out without excessive time and cost.
- A large number of metroligically different CMMs can be simulated.
- Several simulated artifacts can be used without needing explicit separate calibration.
- Parameters and Influence factors can be isolated and varied with great control allowing the software testing to be specific in its focus.
- It is easy to obtain a quantitative measure of the extent to which a reported uncertainty is under or over valued.
Method 3 – CVE  (cont.)

• **Disadvantages:**
  - Well-understood models are readily available for only some influence factors and parameters.
  - Computer simulated measurement situations do not include all real world effects. (Since certain parameters can be isolated and examined using this method, this can be both an advantage and disadvantage.)
  - The method requires some means to exchange information with the software under test.
Conclusion

• Your day at work that started bad turned into a lengthy, yet interesting project

• You get out of being assigned other work by constantly talking up how difficult and lengthy this problem is

• You end up going to Japan to give a talk on the subject and have a great time doing so 😊