AN ASSESSMENT OF A STANDARD DATA VALIDATION APPROACH COMPARED TO FALCONEER'S PATENTED METHOD

Daniel Chester
Associate Chair
Computer and Information Sciences
University of Delaware
chester@cis.udel.edu

The standard approach to data validation developed or used by many is to find the set of values for the process variables that satisfy all the mass and energy balances (and other constraints) and that are as close as possible to the measured values. FALCONEER uses these balances (and other constraints) in a quite different way. Most notably, FALCONEER constructs secondary models, which are additional constraints that help to determine whether a particular process variable is significantly different from its measured or assumed value. FALCONEER's approach is based on logical inference from patterns of model deviations, while the standard approach is based on numerical optimization in a space with high dimensionality. Here are some concerns about the standard approach which FALCONEER addresses.

The standard data validation approach is based on the fact that a process plant was designed so that all the constraints can be satisfied at the same time, that is, there are numerical values that can be substituted for the process variables and make all the equations balance. There are many sets of such values, since the plant can be operated under many conditions that are considered normal. Measurements, on the other hand, may be inaccurate, and so when their values are plugged in (along with the assumed values for unmeasured variables), the equations don't balance. When one or more of the measured values is an extreme distance from the true value, the solution that optimization finds is going to be skewed toward the bad values; all the other variables are going to be off from their measured values in an attempt, so to speak, to distribute the error over all the variables as evenly as possible.

Thus all variables are likely to be distorted to some extent. If the models aren't too nonlinear, this distortion might be small and the method will undoubtedly have some benefit, but I still don't trust the values to be as accurate as they are claimed to be. Optimization in high-dimensional spaces (i.e., dimensions of 2 and higher) is notoriously hazardous, since there are often many local optima and it is difficult to find the global optimum. The local optima can be widely scattered so that the solutions found from slightly different starting conditions may be very different from each other. When FALCONEER forms a secondary model by eliminating a variable of interest from two primary models, it expects the secondary equation to be balanced when the variable of interest is faulty (and the other variables are reasonably close to their true values). Examining each variable separately avoids the difficulties that arise with increasing dimensionality of the variable space, i.e. with increasing number of variables included in the modeled process.

Another side effect that might happen when error is distributed over many variables is that rare events may be overlooked. A measured variable that normally is very accurate may deviate significantly from its true value, but the deviation is still fairly small, small enough to be accounted for by tweaking the other variables a little bit without raising any alarms. By only considering a few constraints that are relevant to a particular fault, FALCONEER can isolate the faulty variable so that its error won't be hidden, making the discovery of rare events more likely.

One final point: it is not clear from what is published whether the derivatives of process variables are allowed in their constraints using the standard data validation approach. We discovered over 25 years ago that it is important to have differential equations as models so that dynamic changes can be tracked and deviations from model behavior can be detected more quickly and accurately. If they do allow derivatives, that is good, but if they don't, that would significantly limit the applications that their approach can handle. FALCONEER has always expected that some models may contain derivatives, i.e., rates of change in levels of tanks, etc.

FALCONEER Technologies LLC
4455 Transit Road Suite 3C, Williamsville, NY 14221
Phone: 716 204-0273 Fax: 716 558-7250
Email: falconeer@falconeertech.com