Building Models in the Presence of Measurement Error

Authors: Philip Yates

Multiple linear regression is a common analysis procedure taught in our universities and used in our industry. Engineers are routinely confronted with the task of understanding whether or not a correlation exists between a set of process inputs and a process output. If a significant correlation exists the engineer's attention may then shift to the strength of the measured association. But, does the typical engineer realize that one of the classical assumptions in developing a linear regression model is that the process inputs, or independent regressors, are measured without error? And, that most regression platforms in commercial software implement algorithms that make use of this assumption? What is the impact of such an assumption? Is there a way to build regression models where the assumption of "perfectly observed" independent terms is relaxed? Can the engineer incorporate his knowledge of the measurement tool's precision into the model building process!

? The purpose of this case study is to attempt to address these questions and outline some of the associated difficulties and strategies for dealing with "errors in variables" regression.

The need to consider such an approach should be evident when one considers the challenges that face our industry today. Some examples include: metrology errors consume an ever growing portion of the process operating window, narrow operating regimes limit the size of the design space the engineer can explore, empirical models are routinely developed to implement process control loops, and the number and duration of learning cycles can be impacted by the surreptitious role measurement error plays in our decision processes.