Originally appeared in HYDROCARBON PROCESSING® March 2007, pg 118. Posted with permission.

Y. ZAK FRIEDMAN, CONTRIBUTING EDITOR

Zak@petrocontrol.com

Inferential models that correlate but do not predict

A friend has asked me to help solve a dilemma. A neural net model developed in his plant had a near perfect fit against lab values but in spite of that the model's prediction ability was poor. I was sleeping on this for many nights trying to come up with a plausible explanation, and then one morning I found this on Yahoo financial news: "Researchers have found that the healthier you are, the richer you'll be. Find out why a robust body often leads to a strong bank account." Not a joke but a serious headline. The possibility that wealthy people could be healthier because they can afford a healthier environment and better health care is not even under consideration.

It dawned on me that regressed inferential models are by and large developed from steady-state data. The steady-state observer sees that wealthy people are healthier, but there is no guarantee that the correlation machine can discriminate between cause and effect.

Plant examples. Here is one taken from the world of advanced control: "On our gasoline reformer, reducing the reformate stabilizer bottom temperature results in lower reformate octane." I was dumbfounded when I first heard that assertion. How is it possible that reducing the effectiveness of stabilization would have anything to do with octane? Octane is determined in the reactor, not in the downstream distillation columns.

But there was a process engineering explanation. When reformer feed is too light, the reforming reaction becomes less effective, and reformate octane declines. Reformer feed is cut in the crude unit, whereas stabilizer bottom temperature reduction is the result of light reformer feed, not the cause of low reformate octane. While the stabilizer bottom temperature is mostly dictated by the feed boiling point, it can also be changed to a limited degree by changing the C_4 separation, which has absolutely no bearing on octane. Thus, the inference of octane as a function of stabilizer bottom temperature fits the data surprisingly well, but misses the boat on predictability. The octane number inference would be better off using crude unit data to predict reformer feed boiling point and using that as an input to the octane inference.

How do you like the following assessment about the relation between crude unit side draw and internal reflux? "On our crude unit, increasing the kero draw causes the overflash flow to also increase." That from an APC engineer who did not realize his inference model stands in contradiction to mass balance principles. Upon further investigation the story unfolded: During operation with sweet crudes the kero draw is maximized. Another unrelated sweet crude operation requirement is that the lowest draw—atmospheric gasoil—be taken into the diesel pool instead of the FCC. When processing sweet crude, operators increase the overflash flow to improve the atmospheric gasoil qualities. To the steady-state regression machine it looked like kero maximization was the cause of increased overflash.

Engineering judgment needed. There have been arguments that computers are getting better and pretty soon we will be able to apply dynamic regression, eliminating the need for engineers to discriminate between cause and effect. I find such arguments naïve. We have to know ahead of time what the root causes are and how we can measure them. If causes cannot be measured quickly that would confuse the regression machine again. In the reformer example, the cause is measurable in a different unit, and to measure this cause one has to install a reformer feed boiling point inferential model. By the way, there ought to be a knowledgeable engineer around to say that reformer feed is cut in the crude unit and that affects reformate octane. Whichever way you approach this subject you come to the conclusion that without solid knowledge of the unit, and engineering judgment in determining the independent input variables, it would be hopeless to attempt the inference model.

I, and others,[1, 2, 3] have argued that no matter what inferential methodology one chooses, the developer of that inferential model must be thoroughly familiar with the unit—indeed, not only with the unit of interest but also with upstream and downstream units. Looking at inferential modeling with process engineering eyes, I never thought about confusing cause and effect, but about the ability to choose a set of inputs that "has the inferential information in it," and that those inputs, to the extent possible, not be dependent on each other, because Gaussian theory requires the independent variables to be independent. Further, I thought that the inferential modeling engineer, being thoroughly familiar with the unit, would be in a position to judge whether the regression coefficients are reasonable.

I see now that I missed an important reason as to why only knowledgeable people should develop inferential models. There is no statistical technique to detect false correlation coefficients with no predictability in them. Only a decent process engineer could do that. **HP**

LITERATURE CITED

1 Friedman, Y. Z., "More about inferential control models," HPIn Control, *Hydrocarbon Processing,* February 2005.
2 Friedman, Y. Z., "Choosing inferential modeling tools," HPIn Control, *Hydrocarbon Processing,* January 2006.
3 King, M. J., "How to lose money with inferential properties," *Hydrocarbon Processing,* October 2004.

Y. Zak Friedman is a principal consultant in advanced process control and online optimization with Petrocontrol. He specializes in the use of first-principles models for inferential process control and has developed a number of distillation and reactor models. Dr. Friedman's experience spans over 30 years in the hydrocarbon industry, working with Exxon Research and Engineering, KBC Advanced Technology and since 1992 with Petrocontrol. He holds a BS degree from the Israel Institute of Technology (Technion) and a PhD degree from Purdue University.