Originally appeared in HYDROCARBON PROCESSING® December 2007, pg 134. Posted with permission.

Y. ZAK FRIEDMAN, CONTRIBUTING EDITOR

Zak@petrocontrol.com

Inferential controls that correlate but do not predict

I recently received a letter from Pavlos Ioakimidis of Hellenic Petroleum responding to my editorial about inferential models that correlate but do not predict.[1] That March editorial argued that regression-based inferentials do not necessarily discriminate between cause and effect, and hence, may be improper to use. Upon reading the letter I realized that my statement was not accurate and should be modified. Pavlos' letter (with minor editing modifications) follows.

"Your argument was that inferentials that are not based on the actual root causes of a physical property, eventually fail to predict. To visualize that you presented the example of reformate RON that shouldn't be correlated with a relevant reformate debutanizer tower bottom temperature, as the latter is probably the effect of a change in actual unit feed quality that in reality concurrently altered RON and this temperature. You proposed the use of feed boiling point as a more appropriate input to this inference.

I clearly disagree with your statement. Perhaps debutanizer bottom temperature is not a good candidate input to a RON inference because it is affected by bottom C_4 content (mentioned in the article), but not because it cannot affect the reformate RON.

My best rules for selecting inference inputs follow.

- *The candidate input should have a one-to-one correlation with the inference of interest, or, to the extent the input is affected by other variables, we should at least be able to compensate for those side effects (like this temperature you mention, or the reformer reactor DTs).*
- *The inputs should be independent of each other (you mentioned later in the article).*
- *The inputs should be measured accurately, reliably and conveniently.*

My main point is: cause and effect relations are irrelevant to success of inferential model development."

My comments. "Guys who want to be tall should play much basketball," is an inferential model, but if you go on to play basketball in a plight to become taller—that inference would fail. Mr. Ioakimidis has a point in saying that root cause or not, while the control strategy is flawed, the inference of basketball players being indeed tall is accurate. The problem of confusing cause and effect comes into play only in selecting manipulated variables. In the personal height control case, playing basketball is not a valid manipulated variable. In the reformate octane control case, debutanizer bottom temperature is not a valid manipulated variable.

I would like then to modify my argument about selecting inferential inputs, remembering that our main purpose is inferential control and not merely theoretical modeling. There are two kinds of inferential inputs: basic and secondary. The basic and most important kind must respond directly to root causes. Of course manipulated variables themselves, which are by definition root causes, could serve as inferential inputs of the basic kind. The second class of nonroot-cause inferential inputs are not valid to serve as manipulated variables and their main purpose is to improve model accuracy. In our reformer example, extent of reaction is a basic input. Feed properties could also be basic inputs, and we can specify manipulation of feed properties, for example cut range, to control reformate octane. But in practice such a mechanism does not normally exist—feed properties are secondary inputs. Extent of reaction is in direct response to reactor temperatures, whereas feed properties are not.

Whether debutanizer bottom temperature can be input in lieu of feed distillation range is another consideration, and it depends on our ability to correct that temperature for the bottom butane content. In my view it would be best to have a model based on upstream distillation column conditions, for example, infer reformer feed cutpoint from crude column conditions, and front cutpoint from naphtha splitter conditions.

If you do not have upstream cut models, would it be OK to use pressure-compensated debutanizer bottom temperature as an input that relates to the feed boiling range? That would be somewhat problematic. You would need to correct for the influence of bottom C_4, and even after that, the bottom temperature is a bubble point and does not precisely reflect the reformer feed boiling range. Either way, bottom temperature is not to be used as a manipulated variable for octane control.

Following this largely theoretical discussion, one could get the impression that reformate octane is much affected by the feed distillation range. That is true only to a limited extent. Feed PNA (paraffin, naphthene, aromatic) composition is a far more important inferential input. Many reformate octane models fail because PNA is not known. I have in the past inferred PNA from unit measurements, and that makes it a secondary input variable, which serves to improve model accuracy but would not be used as a manipulated variable.

Having corrected my cause-and-effect statement, I would like next to consider Pavlos' three rules for selecting inferential inputs: correlation against the property to be predicted, independence and ease of measurement. The discussion above has added a fourth rule dictating that at least some of the inputs must respond to manipulated variables. In my view that is still not enough and a fifth rule is needed: the set of inputs must "have the total information in it." Should an input set be incomplete, the inference would drift, resulting in poor reliability and requiring frequent bias updates. In the reformer model the feed property information serves to prevent inferential drifts. Another way to look at it is: APC would obtain the most value from an octane model that responds to feed property changes, such as during crude or mode switches. At other times the unit operates roughly at steady state, and then daily lab correction is enough to keep the unit running near its target octane. **HP**

LITERATURE CITED

[1] Friedman, Y. Z., "Inferential models that correlate but do not predict," HPIn Control, *Hydrocarbon Processing,* March 2007.

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.