Originally appeared in HYDROCARBON PROCESSING® January 2006, pg 15-16. Posted with permission.

Y. ZAK FRIEDMAN, CONTRIBUTING EDITOR

Zak@petrocontrol.com

Choosing inferential modeling tools

Goel and Shah of Reliance Industries in India published a recent article, claiming that spreadsheet regression is the preferred choice for inferential modeling based on price considerations.[1] Since concepts of "cheap" and "expensive" correlate with labor rates, that claim may be valid in India, but what about the rest of us?

Before delving into this question, some comments on the accuracy of the claim are due.

Simulation versus first-principles inferentials. Goel and Shah assert that first-principles inferentials methodology involves rigorous simulation. That could be true, in the sense that rigorous simulation is based on first principles. However, the complex problem of inferring via the use of simulation has not in general been solved. To my knowledge, only one such case was reported,[2] and even then the simulation was used for infrequent biasing of a simplified engineering model, not for minute-by-minute inference.

That, in the end, is only a semantic argument. But it is good to clarify that, by and large, first-principles inferences employ simplified engineering models—using chemical engineering and process engineering concepts—and are not heavy on computation, whereas simulation-based inferential models practically do not exist.

Skill level needed for developing inferential models. Goel and Shah assert that the skill level requirement for developing inferential models is high for first-principles models but low for regression models. Most people in the inferential models business would take issue with that assessment. Selecting a set of inputs to an inference model is an art form, and it is likely that the uninitiated would fail miserably.

Regression models are based on Gaussian rules, which require all inputs to be independent. Such independence, however, does not exist on a real unit because measurements are connected by mass balance, energy balance, equilibrium laws, etc. Typically all temperatures on a distillation column trend up and down together. Regressions based on correlated input data are apt to yield incorrect models, and that is where the expert would make a difference. He/she would not choose simple measurements as inputs but would mathematically group the inputs in an effort to minimize interdependence. This thinking process certainly involves knowledge of first principles.

Another area where an expert would make a big difference is converting models to work in a dynamic control situation. All regression inferentials are based on steady-state data in open loop. They often go unstable in closed loop because their inputs are out of phase. And they err during transients, often to the point of being turned off during unit disturbances, at the time when APC is needed the most. An expert modeler should be able to use predictive techniques plus knowledge of the unit to bring all inputs into phase and make the model work at all times.

Further, use of daily lab data is problematic. Daily lab data contain many errors, mainly because of sample timing and handling issues, and it takes a real expert to identify unreliable lab data. Simply taking lab points that do not fit the model and labeling them "outliers" does not improve the model quality.

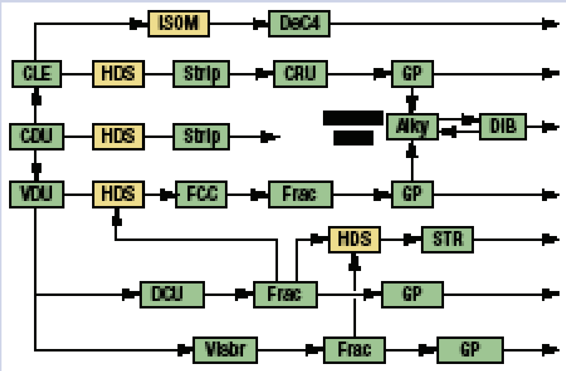

FIG. 1 Several inference models (shown in green) are available for refinery processes.

Comparing neural networks to regression. Both regression and neural networks use Gaussian rules and suffer equally from the problem of dependent input variables. Neural net packages cost a premium; what does one get for this money beyond a good user interface? Let's view the differences:

- Neural nets continuously come up with new terms, linear and nonlinear, until they obtain a perfect fit of the model against the lab data. But such a fit is not desirable. It is well documented that, as the fit against calibration data gets better, prediction of future values gets worse, and one (an expert?) should stop the identification procedure before reaching the bad predictability point.
- Neural nets do not, in general, let you view model coefficients. That is not a good feature: because one way to determine model reliability is to regress half the data, then the other half, and compare the coefficients of these two models.

Based on these two main differences, one would tend to agree with Goel and Shah that a spreadsheet regression is no worse than neural net. I would welcome a comment from the vendors who offer neural network packages, elaborating the advantages of neural networks if they exist.

Comparing first-principles models to regression. My February column[3] highlighted the six main advantages of first-principles models over empirical models:

1. Regression requires independent inputs, which do not generally exist, and first-principles work—or at least knowledge is required—to approach input independence.
2. Empirical models require large amounts of lab data and,

hence, they must make use of poor-quality daily lab data.

3. Empirical models must identify a large number of coefficients. The more coefficients, the better the calibration fit, but the poorer the predictability.

4. First-principles models provide the means for checking instrument errors. Empirical models calibrated using erroneous inputs will not work.

5. There is no replacement for process engineering in terms of predicting equipment performance, i.e., predicting product qualities from equipment measurements.

6. Process modifications necessitate minor changes to a first-principles model, but an empirical model would have to be redeveloped from scratch.

Thus, the question is not whether first-principles models are better—they are better by definition—but whether they exist. In both cases, we need an expert to make the models work well, but for a first-principles model to exist, the modeling knowledge must be in a packaged mathematical format that can be programmed into a computer. If the knowledge exists but is not yet packaged, development and validation would be expensive. Where models do not exist, we must resort to empirical models, and develop them with the aid of an expert who knows not only chemical engineering principles, but also the practical operation of the unit.

Choosing inferential modeling technology. Now to the initial question of which model would serve us best, considering both quality and price. I acknowledge that my business is supplying first-principles models, but hope that the facts speak for themselves regardless of the conflict of interest.

I find it difficult to endorse neural network models over regression, and would therefore make a judgment only on first-principles versus empirical models. The theoretical advantages of first-principles models suggest that, where a first-principles model exists for a reasonable price, it would be a first choice. Existence here means: The model principles are understood, and the model has been packaged and validated in a similar unit. A model that has not yet been packaged would have high development costs, and the user may accept being a beta site, but should also be prepared with a fallback regression model. Where a model does not exist, the modeler must resort to regression, but not without fully understanding the unit relations and coming up with a set of inputs as close to being independent as possible.

What first-principles inferential models have been packaged to the point they can be trusted to give good results? To no one's surprise, distillation processes are very well understood, and good first-principles models exist for both main fractionators and simpler distillation columns. Many papers have been written about such models demonstrating that, by and large, the models work. The literature cited lists two recent articles as examples.[4,5] Given that many product properties are controlled by distillation columns, most inferential needs have actually been covered by first-principles models for years.

And what about reaction processes? We know of the following packaged first-principles reactor models and are inviting readers' comments about other reactors.

- Reformer: Estimates reformate octane or aromatic content
- Alkylation stoichiometry: Estimates isobutene/olefin ratio
- FCC: Estimates catalyst circulation, coke make and feed properties
- Coker: Estimates drum fill rate
- Visbreaker: Estimates visbreakate stability.

Fig. 1 presents a simplified refinery unit diagram and highlights available first-principles inferential models in green, illustrating the extent of penetration of such models. Hydrotreating is a glaring, unfulfilled necessity. Unfortunately, severe hydrotreating does not lend itself to inferential modeling by any technique because the model would require detailed knowledge of the feed sulfur molecular structure, which does not exist. We would either have to conservatively overtreat or resort to expensive ppm-level sulfur analyzers. The other missing models, for isomerization and hydrocracking processes, are probably missing not because the reactors are not understood, but because no one has tried. **HP**

LITERATURE CITED

1 Goel, S. and J. Shah, "Simple techniques for generating process control inferentials," *Hydrocarbon Processing,* October 2005, pp. 81–84.

2 Schuler, M. G., Y. Z. Friedman, M. G. Kesler and P. Belanger, "Refinery uses column data to infer and control product properties," *Oil & Gas Journal,* Feb. 19, 2001.

3 Friedman, Y. Z., "More about inferential control models," "HPIn Control," *Hydrocarbon Processing,* February 2005, pp. 17–18.

4 Adnan, A., N. Md. Sani, S. Y. Nam and Y. Z. Friedman, "The use of first-principles inference models for crude switching control," *PTQ,* Fall 2004.

5 Friedman, Y. Z. and M. G. Schuler, "First-principles inference model improves deisobutanizer column control," *Hydrocarbon Processing,* March 2003, pp. 43–47.

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 more than 30 years in the hydrocarbon industry, working with Exxon Research and Engineering, KBC Advanced Technology and in the past 12 years with Petrocontrol. He holds a BS degree from the Israel Institute of Technology (Technion) and a PhD degree from Purdue University.