

The Other 21 People on the Team

Where are your advanced modeling techniques most likely to fail?

Wayne S.Obetz, Ph.D.

The Premier Business
Leadership Series

October 2008

Key Points

- Time and attention flow to newer, more desired things at the expense of older, less glamorous things
- The people responsible for the routine updating of your data are as critical to the success of your modeling efforts as anyone else
- The most advanced models depend on the basics of data validity and integrity, but this concern often takes a back seat to other concerns
- Keeping your models as simple as possible reduces your exposure to problems with data acquisition and validity

We tend to spend time on what we desire, not necessarily on what we need

- Automotive engineers spent the three decades since the previous energy crisis improving luxury and power



Eventually we learn where our time and effort would have been better rewarded

- Maybe they should have been spending their time improving fuel economy



Gas lines in 1979

Gas prices Summer 2008



We tend to celebrate the new and flashy, and have less enthusiasm as these items age and dim

- Politicians and the general public have an appetite for ribbon-cutting ceremonies when a building or bridge is constructed, but they generally don't celebrate the routine maintenance of these structures



How many of the 22 people on a football field have possession of the football at any one time?



What are the other 21 people doing?

The most advanced models depend on the basics of data validity and integrity, but this concern is often minimized

- We staff modeling projects with the best talent we can find—are we doing the same with those parts of the organization responsible for maintaining the data that feeds the models?



We recognize the need to put resources into our advanced modeling efforts

- How complex is your process for vendor selection when doing advanced modeling?
- How difficult is it to justify expenditures on new software with more capabilities than your current generation of software?
- How many staff members and consultants do you have working on these advanced models?

How many resources are dedicated to maintaining your data validity, availability, and alignment?

- Are your production datasets being validated every production cycle?
 - Are you validating that a dataset was written, or are you establishing that certain values have been properly written to the proper location?
- Are your production datasets being updated dataset-by-dataset, or are they updated simultaneously?
 - Are all updates being done at once, or do your users need to check file creation dates to be sure that all files have been updated?
- Do your weekly and monthly ranges have the same meaning across all your datasets?
 - Do you have multiple providers of data and are all their date ranges in synch?

The most advanced models depend on the basics of data validity and integrity, but we tend to minimize these concerns

- Systems are geared towards the generation and acquisition of data, but not its validation
 - Generally designed to make the transaction in the field easier, with less concern for data validity
- Collect all available data, even that which you have no current plans to analyze
 - Justified by reasoning that you cannot retroactively collect this data once you do figure out what you want to do with it
 - Cannot validate this data until this decision is made
- Pilot testing of large systems needs to involve all the people who will eventually be asked to use it

How many datasets are you updating on a regular basis?

- An inventory of datasets for which my group is currently responsible shows 331 separate files that must be kept up-to-date
 - 38 of these files are updated weekly, with 14,736 individual variables
 - 226 files are updated monthly, with 35,883 individual variables
- Some datasets are updated locally, some are updated by an off-shore team
 - The level of effort expended on validation of these datasets varies considerably
 - Those datasets updated by end-users tend to be more thoroughly scrutinized

How many variables are required to update your forecasts and analytical models?

- A typical forecast in one of my previous roles required the monthly update of 24 variables, from no fewer than eight separate sources, to generate the predicted sales in one of our three sales channels
 - Data came from internal sources (Finance, manufacturing, customer relations, marketing teams, and sales teams), syndicated sources, external purchasing groups, and a specialty forecasting vendor
- Forecasts for the remaining two channels were generated by two other groups, each responsible for updating their own multi-source models
- Forecast value added analysis showed that for one-quarter of the product lines, an exponentially-smoothed time-series model of total shipments consistently outperformed the more complex model in place
- The same analysis showed that for 17 of 20 products, exponentially-smoothed time-series models consistently outperformed the more complex model for the sales channel for which we were primarily responsible

Why the salmon I hooked is still in the river

After three years of trying, I finally hooked a salmon on my fly rod. In the process of fighting the mammoth creature of the sea, which component of my fishing gear failed me?

- a) the 2006 Field & Stream Best of the Best award-winning carbon fiber fly rod, made with an exclusive thermoset thermoplastic resin
- b) the third generation fly line, made with the slickest coating available, with its gel-spun backing (60% stronger than standard Dacron)
- c) the large arbor reel, capable of retrieving line more than three times faster than a standard reel
- d) the high-performance waders with the hi-grip, no-slip soles
- e) the mostly overlooked 2¢ screw holding the reel handle to the spool

Questions and Comments

Wayne S. Obetz, Ph.D.
Senior Manager
Quantitative Commercial Insight

AstraZeneca
1800 Concord Pike
P.O. Box 15437
Wilmington, DE 19850
wayne.obetz@astrazeneca.com
wobetz@gmail.com