Academic Life, Citation Counts and Statistical Models

Luncheon Presentation at the 46th Annual Fall Technical Conference
Valley Forge, PA
October 18, 2002

William H. Woodall
Virginia Tech
Blacksburg, VA 24061-0439
bwoodall@vt.edu

Introduction

I gave considerable thought to my topic for today, given an audience of many who have just eaten too much and stayed up too late last night. I will try to be interesting, relevant and, most importantly, brief.

I will discuss the academic reward system, the use of citation counts to measure the influence of statistical ideas, and the use of statistical modeling.

Publishing and Academia

In academia it is well known that faculty members are primarily rewarded for research and grant support, not teaching. The most important rule of academic life is not "Publish or Perish". This is far too strong. Before tenure it is "Publish or Vanish" and after tenure it is "Publish or Languish." Some rewards of publishing and grant support include less teaching, smaller classes, better students, graduate student assistance, more money, more travel, increased status and universal respect (?).

In academia, publication of ideas in refereed journals has been what counts, not necessarily their practicality or acceptance in statistical practice, at least in the short run. Also, review papers, case studies, interdisciplinary papers and what some call negative research have been heavily discounted. Decisions are typically made on the number of publications and the journals in which the publications were published.

There are many statistical journals in which to publish ideas. *Journal Citation Reports* lists 70 in the Statistics and Probability category. Some are much more highly regarded in academia than others. There are ordered categorical rankings varying by university. ASA, IMS and RSS journals are always rated highly. *Technometrics* is ranked in the same category or higher than the *Journal of Quality Technology* (*JQT*). To some extent the status is positively correlated with the level of mathematics, but there may be more important factors such as citation impact and acceptance rate.

One often hears that *JQT* and *Technometrics* have become too mathematical for practitioners. Journals do have a natural tendency towards becoming more mathematical.

One must keep in mind, however, that some in academia would automatically consider any work in *JQT* and *Technometrics* to be somewhat simple-minded from a mathematical point of view and would tend to discount the value of the ideas for that reason alone. Some academics criticize us for including industrial practitioners on our editorial review boards.

Use of Citation Counts

Each reference listed at the end of a published article is a citation. Journals, Ph.D. programs and individuals are increasingly ranked by the numbers of citations received by articles. This metric is only a rough measure, however, of impact and intellectual influence. Counts do not reflect, for example, unpublished applications by practitioners, citations in textbooks or implementation in software.

The primary index for ranking journals is the average number of citations to papers within two years of their publication. This is too short for the field of statistics and may reflect primarily self-citations. *Statistical Science* had the highest index for 2001, a value of 2.00. Citations of a paper may be delayed for quite a few years. Almost 80% of the citations to *Technometrics* articles in 2001 were to *Technometrics* papers more than ten years old.

Stigler (1994) discussed some of these issues in a study of citation patterns in 33 journals of statistics and probability. *JQT* was not included in his study even though its citation impact would have exceeded that of about 14 of the covered journals. *Journal Citation Reports* now has *JQT* appropriately listed in the Statistics and Probability category, so it should be included in any such future study. *JQT* is a statistical journal. Half the current editorial review board members of *JQT* are Fellows of ASA, higher than the corresponding percentage for *Technometrics*.

The use of citation counts in evaluating the performance of individuals can be misleading. Self-citations, for example, should perhaps be excluded, although some argue that they properly reflect the impact one has on oneself. Support for use of citation counts among academics seems to be strongly positively correlated with the number of citations. Some professors even have citation counts posted at their websites.

The most cited mathematical scientist in the 1990s was David Donoho of Stanford. (Nine of the top fifteen are statisticians, and 15 of the top 25.) His 1994 *Biometrika* paper on wavelets, for example, has been cited 351 times. He has even written an essay on how to get papers cited (www.in-cites.com/scientists/DrDavidDonoho.html). For perspective, the 1951 Box and Wilson *JRSS* paper on RSM, celebrated at last year's FTC, has been cited about 900 times. After about 20 years such citation classics start to receive fewer citations even though their impact persists.

In an upcoming ISI ranking of the 250 most cited mathematical scientists over the period 1981-1999, statisticians are very well represented. George Box and Jeff Wu are included.

Most cited papers in science are cited only once. The percentage not cited at all is unknown. It is not easy to get 100 citations (to reach "centurion" status as referred to by Donoho). The 1992 *Technometrics* panel discussion on Taguchi Methods edited by Vijay Nair, which included a who's who of DOE experts, has just over 100 citations. According to ISI, only one out of 1,000 papers in the mathematical science got cited 133 or more times in 1981-1997 (only one in 10,000 reached 300 citations).

JQT has the following six centurion papers thus far: Nelson (1969) on reliability, Derringer and Suich (1980) on multi-response optimization, Montgomery (1980) on the economic design of control charts, Kackar (1985) on Taguchi methods, Kane (1986) on capability indices, and Montgomery and Mastrangelo (1991) on SPC with autocorrelated data. *JQT* fares pretty well in citation impact comparisons.

The most highly cited papers in *Technometrics* are Wold (1978) on variable selection in multivariate analysis, and Hoerl and Kennard (1970) on ridge regression, with roughly 900 citations each.

Quality Engineering doesn't get cited very much. It is more of a practitioner-oriented journal. It doesn't have a blind review process and it has a 70% acceptance rate. These characteristics would cause it to fall into a low rating category in academia regardless of its impact on practice.

Tom Ryan and I are assembling a list of the 25 most cited papers in statistics. The three highest citation counts for statistical papers (of which we are aware) are the 1958 *JASA* paper by Kaplan and Meier on survival function estimation with over 24,000 citations, the 1972 *JRSS-B* paper by D. R. Cox on proportional hazards modeling with 16,611 citations and the 1955 *Biometrics* paper by D.B. Duncan on his multiple comparison test (written while at Virginia Tech) with nearly 13,000 citations. Each of these numbers exceeds the total number of citations to *all JASA* articles in 2001 (10,812). These three papers are in the top thirty (or so) most cited papers in all of science in 1981-1995.

D.W. Marquardt's 1963 *SIAM Journal* paper on estimation in nonlinear models was fourth with 9,000 citations. As many of you know, Marquardt worked for DuPont and was a leader in the quality movement. He received, for example, the Shewhart Medal.

As other examples, the 1979 *Annals of Statistics* paper on the bootstrap by Efron is number 16 with 1,700 citations. Hoerl and Kennard (1970) and Box and Wilson (1951) are each around number 25.

There is a 1951 paper in the *Journal of Biological Chemistry* with 250,000 citations.

Checking citation counts can be humbling, so let me take this opportunity to ask you to please cite my papers.

Those of us in academia are not always respected for our ideas, whether cited or not. Eric Ziegel stated in a *Technometrics* book review on the proceedings of a meeting I attended in Japan, "Thirty academic statisticians aren't going to solve many practical problems." I am sure he wrestled with the "m". In an article in *Statistical Science*, David Banks stated that academic statisticians were considered by many practitioners to be "pretentious parasites."

One's view regarding the potential impact and usefulness of new statistical ideas depends to a large extent on how one tends to view the models used and the role of models generally.

The Role of Models

What is the role of models (i.e., distributional assumptions and assumptions regarding dependence) in statistics? Breiman (2001) stated that 98% of statisticians fall into the category of relying on models to justify their methods. He does not view this fact positively, by the way, saying it has lead to irrelevant theory, questionable conclusions and lack of involvement in interesting current problems. He believes that if a method seems to work well, we should use it, even if we don't understand how it works.

A genetic algorithm-designed neuro-fuzzy regression tree may work well for classification based on multivariate data, for example, but no one would fully understand why.

In general, journals publishing statistical papers in the quality area have required models.

I think the percentage of industrial statisticians relying on models is much lower than 98%. Included in those who seemed to have little or no use for statistical modeling are some of the most influential people in the quality area, e.g., Genichi Taguchi, Dorian Shainin, and W. Edwards Deming, all Honorary Members of ASQ. Also included to some extent is Mary Lietnaker, who wrote a special publication for the Statistics Division on statistical thinking, and possibly many in the Statistics Division of ASQ. (I can't help but ask here that if you have a special publication, please submit it to *JQT*.)

I strongly support the following concepts of statistical thinking endorsed by the Statistics Division:

1. There is always variation.
2. Work is a series of interrelated processes.
3. Reducing variation corresponds to improving quality.

Note, however, that these concepts would also encompass engineering-control-theory-thinking, Pre-control-thinking, and Taguchi-Method-thinking. I think in many cases we need to take the step to statistical modeling. In my opinion, the Statistics Division's Venn diagram representing sets for Statistical Thinking and Statistical Methods also needs a set representing Statistical Modeling.

This gets to the heart of what is "statistics" and what isn't "statistics"? I agree with Friedman (2001), and Straf (2002) that we need to broaden the field of statistics, but not all data analysis should be considered to be statistics. It would take a book (or more) to define what is and what isn't statistics. I can see eyes in the audience glazing over already at the thought of it.

Some mathematical statisticians seem to believe the models are reality or more important than reality. Thus, they believe, for example, that Shewhart charts should be replaced with CUSUM charts. Pollak (2001) stated, "………..other (methods) remain popular even though sub-optimal (e.g., Shewhart and EWMA control charts)." (Optimality here is very narrowly defined.)

As always, there is a ditch on both sides of the road. I'll take the easy way out for now and simply agree with Box (1979, p. 202) in that "all models are wrong, but some models are useful." This could be the most-quoted quote in statistics! It appears twice just in the most recent issue of *JASA*. The citation, however, is rarely given.

Examples in Industrial Statistics

Jim Lucas once said, "Research in SPC is an inch wide and a mile deep." I've spent many years down in this hole, which, I might add, was quite a bit deeper from Jim's contributions. Now Jim is doing research in DOE and the dimensions of his work are under evaluation. I also spent several years studying fuzzy logic. I would like to include examples about which I am familiar.

Much of my work has been advocating to practitioners what not to do. This may help to explain my level of popularity.

In my view, research in statistics often involves evaluating, clarifying, or improving current methods proposed in the literature or methods used in practice. If the method evaluated turns out to be based on a really, really bad idea, then the research is sometimes branded "negative." Thus, the term "negative" is mostly a function of the quality of the method evaluated. Negative research is discounted by some academics, but a paper exposing the flaws should count at least several times more than a flawed paper. Also, let's face the fact that there have been quite a few flawed quality methods proposed over the years.

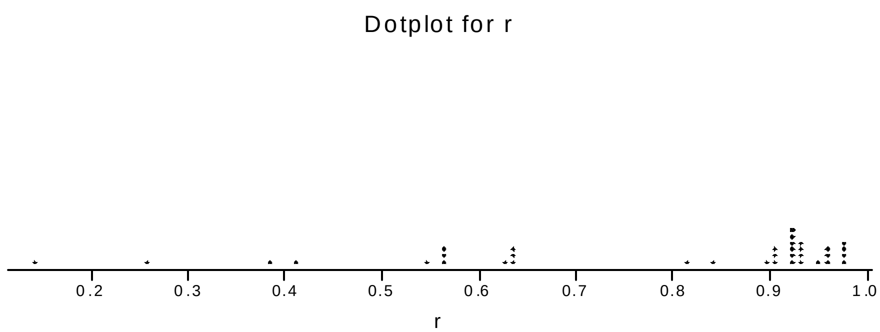
The following are some examples of ideas in industrial statistics with a discussion of the use of statistical models and some citation information:

1. **Economic design of control charts.** With economic design, one builds a cost model and determines the monitoring procedure to minimize the expected cost. A. J. Duncan's 1956 *JASA* paper on economic design has been cited almost 200 times, ranking highly among SPC papers. The approach encompasses about 25%

of all of the research in SPC, but seems, however, to have had very little impact in practice.

I wrote Deming three letters in the early ‘80s with the purpose of trying to get him to “rebuke” those taking the economic approach. No one could rebuke better than Deming. I favored the statistical approach, which to my great disappointment, Deming rejected completely and wrote against in *Out of the Crisis*. I’m fortunate he didn’t mention me specifically as being particularly misguided and lacking profound knowledge.

The economic approach does have problems, although the approach of Saniga (1989) can fix some of them. A recent *JQT* paper proposed the economic design of MEWMA charts for which the value $r = 1$ results in no EWMA feature. The value $r = 0.2$ is standard. The values of r obtained are shown below. Something is obviously wrong here. I think much of economic design is flawed.



In my opinion, research in SPC needs to be directed toward the monitoring of multi-stage processes and product and process profiles.

2. **Taguchi methods.** Most of you are more familiar with Taguchi’s DOE methods and ideas, but he has proposed on-line methods as well. Taguchi provided a gold mine of topics for interesting publications. He was a godsend to the academic community. The most cited papers of a number of very well known industrial statisticians are related to Taguchi’s ideas and methods. Taguchi has had a tremendous impact on industrial statistics, but he won’t show up in any most-cited lists because he wrote books, not refereed journal articles.

   How much impact does Taguchi have? An American Supplier Institute claim is that he has much more influence now than ever. A 2001 *Quality Engineering* paper by a Taguchi co-author was "Taguchi Methods in American Universities and Corporations". Think for a moment about how you would study this. They surveyed only five universities. Two of these were universities where one of the coauthors had worked. One respondent was the person who designed the survey. One of the five universities was in Canada! It is clear to me that Taguchi and his co-authors do not have an understanding of the statistical principles and methods they seek to replace.

Now there is the **Mahalanobis-Taguchi System (MTS).** This is a new set of multivariate methods. Mahalanobis was obviously not directly involved since he passed away 30 years ago. (There is a photo of Mahalanobis, Shewhart, Taguchi, and Prime Minister Nehru taken in 1954.) There is a rejection of probability in MTS. Taguchi writes, "The core of statistics is distribution. In quality engineering, distribution is not considered." There is considerable freedom in developing new statistical methods if one releases oneself from the constraints of statistical models and measures of statistical performance. The MTS methods include the comparison of two treatments based on two individuals (one per treatment). Taguchi foresees clinical trials based on a single person in each treatment group. These ideas run counter to the most basic statistical principles.

Taguchi states in his discussion of our MTS paper, by the way, that since the controversy over his methods has subsided, that statisticians now support the use of his ideas and methods!

3. **Fuzzy Logic.** There are some parallels between the fuzzy logic and Taguchi method phenomena. Both movements have been referred to as cults. Of course they see us as members of the "cult of probability". The entire basis for the existence of *IEEE Transactions on Fuzzy Systems* relies on the fact that we as statisticians supposedly cannot use probability to measure the degree to which a glass of water is potable. Those using probability are said to be limited to yes/no responses while fuzzy logic can be used to give a membership value between 0 and 1 to reflect the degree to which the water is potable.

   Lest you believe fuzzy logic is currently having little impact, *Fuzzy Sets and Systems* had more citations in 2001 than *Technometrics* (3307 vs. 2587). In my view, *FSS* should not be included in the Statistics and Probability category. Lotfi Zadeh's 1965 *Information and Control* paper on fuzzy logic has been cited almost 5,000 times. If included on our list of most-cited statistics paper, Zadeh's paper would rank seventh. In 1998, Eugene Garfield, founder of ISI, stated that he believed there were only 100-200 papers in all of science that had been cited over 5,000 times (www.garfield.library.upenn.edu).

   The founder of fuzzy logic, Zadeh (2000) stated, "… there is no mechanism within the theory (of probability) to represent the meaning of perceptions in a form that lends itself to computation." He refuses to even acknowledge the concept of subjective probability.

The impact of fuzzy logic has been quite large, although in my opinion it is based on flawed reasoning. Fuzzy methods have been proposed as competitors to many statistical methods.

One paper on fuzzy control charts was in the *Journal of Intelligent and Fuzzy Manufacturing* in 2000. They can't have it both ways. How could there be any submissions? I don't like the implication of this use of "intelligent". It is one of my pet peeves, although my wife Mary says I shouldn't keep peeves as pets. If you aren't doing it their way, guess what way you are doing it? These authors stated,

"It (our fuzzy approach) provides a means of avoiding the generation of false alarms due to random variations in the environments and/or measurement instruments or sensors, while successfully detecting all real process faults and also minimizing the time taken to detect such real faults."

This would be the "holy grail" of process monitoring. Their approach is clearly flawed.

We have fuzzy reliability, fuzzy regression, fuzzy forecasting, fuzzy controllers, fuzzy acceptance sampling, fuzzy DOE and fuzzy RSM. For example, see

Lotfi, A., and Howarth, M. (1997), "Experimental Design with Fuzzy Levels", *Journal of **Intelligent** Manufacturing*, 8, pp. 525-532.

Many of these topics are sure sources of papers. It won't get you a lot of citations, but the research is well worth doing.

The most recent issue of *Quality Engineering* contains two fuzzy logic papers.

I believe statistical models are necessary to evaluate all of these and other data analysis ideas and methods. Case study justification alone is not enough.

Note that some approaches are purported to replace statistical methods entirely, e.g., Grey System Theory that seems particularly popular in the People's Republic of China (Julong, 1989).

My rather simplistic view toward modeling took a major turn after reading arguments from Breiman (2001), Hoadley (2001) and others that many "algorithmic" methods are necessary even though modeling cannot be done. Hoadley (2001) argues that some methods, such as logistic regression, work well as long as one forgets about the assumptions. If one forgets the assumptions, one effectively abandons the models.

We should allow this algorithmic exemption to statistical modeling only when modeling can't be done, not when it just won't be done. How do we decide if an algorithmic method is based on a good idea? The MTS was proposed in order "to fill the need for algorithmic methods." This need could be filled by a random sequence of arithmetic operations on the data. I think a good rule of thumb to use is that if a method is to be used, it should be shown that the method performs well under some model or that it performs well when applied to a wide variety of data sets when compared fairly to the top competing approaches. Is this too much to ask? Actually, it may be, but the abandonment of probability and statistical modeling can lead, and has led, to some absurd conclusions.

Conclusions

It is impossible to measure precisely the impact of a statistical idea, although the use of citation counts reveals very useful information. Citation counts, however, can be large for bad ideas, as in the fuzzy logic case; or for ideas that have not substantially affected statistical practice, as in the economic design of control charts. Some claim Duncan's (1955) approach was not sound. Tukey (1991), for example, referred to it as a "distraction" in his history of multiple comparison methods. If one of the most cited statistical papers is based on a flawed approach, then we should clearly not use citation counts without realizing their limitations.

We will always need new, high impact statistical ideas, but not every data analysis idea is something we should claim. Frank Caplan (1990), as the editor of *QE*, argues all methods have their places. I disagree. I also disagree with Gerry Hahn who stated in his 2001 Deming Lecture at the JSM in Atlanta that statisticians needed to get involved in the next fuzzy logic and the next Taguchi Methods. If Taguchi asked me to get involved in MTS (to form the MTWS?), an event about as unlikely as Mahalanobis contacting me, I would have to refuse.

The following are my conclusions:

1. The academic reward system will increasingly be based on citation counts, especially for senior faculty, even though the use of citation counts can be misleading. The reward system can seem very strange. In many cases, what practitioners hope for isn't rewarded in academia. See Kerr (1975).
2. There is no negative research; there are only crummy methods.
3. The ASQ Statistics Division Venn diagram needs a set for statistical modeling. Abandoning probability and the use of models may be necessary in some cases, but it can be dangerous.

Finally, for completeness and fairness, I would like to point out that the workshop proceedings Eric Ziegel criticized resulted from a meeting in Japan I helped to organize, the "Vth International Workshop on ***Intelligent*** Statistical Quality Control."

Some Related References

Adams, B. M., and Woodall, W. H. (1989), "An Analysis of Taguchi's On-Line Process Control Method Under a Random Walk Model," *Technometrics*, 31, pp. 401-413.

Banks, D. (1993), “Is Industrial Statistics Out of Control?” *Statistical Science*, 8, pp. 356-409.

Barrett, J. D., and Woodall, W. H. (1997), “A Probabilistic Alternative to Fuzzy Logic Controllers,” *IIE Transactions - Quality and Reliability Engineering*, 29, pp. 459-467.

Box, G. E. P. (1979), "Robustness in the Strategy of Scientific Model Building" in *Robustness in Statistics*. R.L. Launer and G.N. Wilkinson (ed.), Academic Press, New York.

Breiman, L. (2001), “Statistical Modeling: The Two Cultures”, *Statistical Science*, 16, pp. 199-231.

Caplan, F. (1990), “Editorial – Acceptance at Face Value”, *Quality Engineering*, 4(2), pp. iii-v.

Cox, D. R. (1990), “Role of Models in Statistical Analysis,” *Statistical Science*, 5, pp. 169-174.

El-Shal, S. M., and Morris, A. S. (2000), “A Fuzzy Rule-Based Algorithm to Improve the Performance of Statistical Process Control in Quality Systems”, *Journal of Intelligent and Fuzzy Systems*, 9, pp. 207-223.

Friedman, J. H. (2001), “The Role of Statistics in the Data Revolution.” *International Statistical Review*, 69, pp. 5-10.

Hoadley, B. (2001), “Comment”, *Statistical Science*, 16, pp. 199-231.

Jones, L. A., and Woodall, W. H. (1998), “The Performance of Bootstrap Control Charts,” *Journal of Quality Technology*, 30, pp. 362-375.

Jugulum, R., and Dichter, A. (2001), “Taguchi Methods in American Universities and Corporations”, *Quality Engineering*, 13, pp. 607-621.

Julong, D. (1989), “Introduction to Grey System Theory”, *The Journal of Grey System* 1, pp.1-24.

Kerr, S. (1975), “On the Folly of Rewarding A, While Hoping for B”, *Academy of Management Journal* 18, pp. 769-783.

Kim, K., Mahmoud, M. A., and Woodall, W. H. (2002), "On the Monitoring of Linear Profiles," to appear in the *Journal of Quality Technology*.

Laviolette, M., Seaman, J. W., Jr., Barrett, J. D., and Woodall, W. H. (1995), "A Probabilistic and Statistical View of Fuzzy Methods," (with discussion) *Technometrics*, 37, pp. 249-292.

Leitnaker, M . G. (2000), "Using the Power of Statistical Thinking", Special Publication of the ASQ Statistics Division.

Lin, C. L., Lin, J. L., and Ko, T. C. (2002), "Optimization of the EDM Process Based on Orthogonal Array with Fuzzy Logic and Grey Relational Analysis Method, *International Journal of Advanced Manufacturing Technology*, 19, pp. 221-227.

Linderman, K., and Love, T. E. (2000), "Economic and Economic Statistical Designs for MEWMA Control Charts", *Journal of Quality Technology*, 32, pp. 410-417.

Lotfi, A., and Howarth, M. (1997), "Experimental Design with Fuzzy Levels", *Journal of Intelligent Manufacturing*, 8, pp. 525-532.

Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*. 4th Edition, John Wiley & Sons, New York, NY.

Nair, V., Hansen, M., and Shi, J. (2000). "Statistics in Advanced Manufacturing". *Journal of the American Statistical Association* 95, pp. 1002-1005.

Pignatiello, J. J., Jr., and Ramberg, J. S. (1991-1992), "Top Ten Triumphs and Tragedies of Genichi Taguchi", *Quality Engineering* 4, pp.211-225.

Pollack, M. (2001), "Comments", *Statistica Sinica*, 11, pp. 368-370.

Ryan, T. P. (2000). *Statistical Methods for Quality Improvement*. 2nd Edition, John Wiley & Sons, New York, NY.

Saniga, E.M. (1989),"Economic Statistical Control Chart Designs With an Application to X-bar and R Charts", *Technometrics*, 31, pp. 313-320.

Stigler, S. M. (1994), "Citation Patterns in the Journals of Statistics and Probability", *Statistical Science*, 9, pp. 94-108.

Straf, M. (2002), "Statistics: The Next Generation", ASA Presidential Address, Joint Statistical Meetings, New York, NY.

Tukey, J. W. (1991), "The Philosophy of Multiple Comparisons," *Statistical Science*, 6, pp. 100-116.

Woodall, W. H. (1987), "Conflicts Between Deming's Philosophy and the Economic Design of Control Charts," *Frontiers in Statistical Quality Control 3*, eds. Lenz, H. - J., Wetherill, G. B., and Wilrich, P. -Th., Heidelberg, Germany: Physica-Verlag, pp. 242-248.

Woodall, W. H. (2000),“Controversies and Contradictions in Statistical Process Control,” (with discussion), *Journal of Quality Technology,* 32, pp. 341-378.

Woodall, W. H., and Davis, R. E. (1993), Letter to the Editor, *IEEE Transactions on Fuzzy Systems*, 2, pp. 43-45.

Woodall, W.H., Koudelik, R., Tsui, K.-L., Kim, S.B., Stoumbos, Z.G., and Carvounis, C.P. (2002), “A Review and Analysis of the Mahalanobis-Taguchi System,” to appear in *Technometrics*. (Discussants include R. Jugulum, G. Taguchi, S. Taguchi, Douglas M. Hawkins and Bovas Abraham)

Xie, H., Lee, Y. C., Mahajan, R. L., and Su, R. (1994), “Process Optimization Using a Fuzzy Logic Response Surface Method”, *IEEE Transactions on Components Packaging and Manufacturing Technology Part A*, 17, pp. 202-211.

Zadeh, L. A. (2000), “A Challenge to Bayesians”, posted at newsgroup comp.ai.fuzzy.

Ziegel, E. (1999), “Review of *Frontiers in Statistical Quality Control 5* edited by H. J. Lenz and P. -T. Wilrich”, *Technometrics*, 41, p. 179.