Best Practices in Modeling and Simulation; Multi-Community Benchmarking

Steve Roemerman, John Volpi, Randal Allen
Lone Star Analysis
Addison, TX
sroemerman@lone-star.com, jvolpi@lone-star.com, rallen@lone-star.com

ABSTRACT

Modeling, Simulation, and Analysis (MS&A) supports a wide range of economic, academic and governmental efforts. Many MS&A communities have agreed on methodology within their field. However, there is little interaction among communities. As a result, best practices in one MS&A community may be unfamiliar to others. This paper describes the Modeling Best Practices Benchmarking Project – an effort to identify modeling practices among professionals who might not otherwise gain insight outside their own communities. Practitioners from many disciplines volunteered to describe their practices and learn from others. The goals were to understand best practices across industries and disciplines and define best practices as a set of standards which apply broadly.

From the interview and survey topics, we developed a check list of 14 best practices for those doing MS&A. We also developed three other check lists of risk factors for some specific MS&A topics. Eventually we identified four best practitioners, all of whom impressively addressed the 14 best practices. Two agreed to be named in our work: The U.K. Metrology Office, and the U.S. Energy Information Agency.

The paper will present a sample of the results along with the best practices identified. An example of the results: most respondents did not know regulatory or statutory standards applicable to their work; most did not use processes important to high integrity MS&A. This is a contrast to exemplars.

ABOUT THE AUTHORS

Steven Roemerman is the Chairman / CEO of Lone Star Analysis since 2004. In addition to previous CEO roles, he has served as an officer or director in more than a dozen corporations in technology, aerospace, finance, non-profits, and transportation. His professional career began at Texas Instruments as a technology systems analyst. He was named a Vice President, VP of Strategy. He joined Raytheon upon the acquisition of TI's systems business. His education includes graduating Magna Cum Laude from the Missouri University of Science and Technology and graduate school at Southern Methodist University. He was recognized for leadership on JSOW, which won an Aviation Week Laurels award, and was named Senior Member of the IEEE. He holds several patents.

John Volpi is the Chief Technology Officer (CTO) of Lone Star Analysis and has served in this role since 2004. He is responsible for all technical activities, Intellectual property evaluation, and process development. He began his professional career at Texas Instruments as a theoretical systems analyst. He was named a Senior Member of the Technical Staff. He later became a General Manager for Active Cellular Antennas at Crosspan Technologies, a wholly owned subsidiary of Raytheon. He holds a B.S. Physics (Illinois Institute of Technology) and an M.S. Physics (Michigan State University). He has over 30 patents awarded or pending. He is a Senior Member of IEEE. In 2012, he was awarded the Tech Titans Award for Corporate CTO by the DFW Metroplex Technology Business Counsel.

Randal Allen is the Chief Scientist of Lone Star Analysis. He is responsible for applied research and technology development across a wide range of MS&A disciplines. He holds a CMSP with NTSA. He is co-author of the textbook, "Simulation of Dynamic Systems with MATLAB and Simulink." He is an Associate Fellow of AIAA. He holds a Ph.D. in Mechanical Engineering (University of Central Florida), an Engineer's Degree in Aeronautical and Astronautical Engineering (Stanford University), an M.S. in Applied Mathematics and a B.S. in Engineering Physics (University of Illinois, Urbana-Champaign). He serves as an Adjunct Professor in the MAE department at UCF.

Best Practices in Modeling and Simulation; Multi-Community Benchmarking

Steve Roemerman, John Volpi, Randal Allen
Lone Star Analysis
Addison, TX
sroemerman@lone-star.com, jvolpi@lone-star.com, rallen@lone-star.com

BACKGROUND

In 2013 and 2014, the authors were developing training materials on MS&A and studying failed MS&A efforts[1]. This led to the questions, "What is best practice?" and "Who are the best practitioners?" We were surprised to find few publications on generic MS&A. We eventually concluded a benchmarking project[2] might answer our questions. We worked with several organizations who promoted the benchmarking and helped us collect data.[3]

Data collection was based on two paths: surveys and interviews.

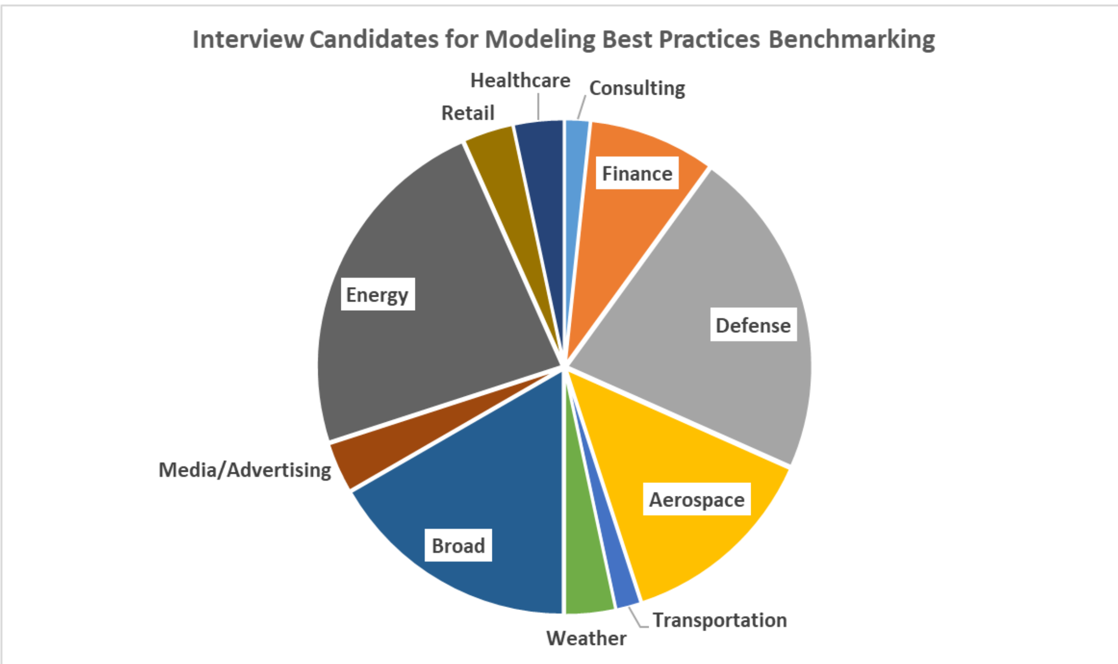

Figure 1 – Interview Candidates

Surveys were the first line of inquiry. We constructed two survey instruments hoping to collect data from MS&A practitioners around the world.[4] The first instrument was short, and was aimed at helping us understand what topics would be of interest. We also hoped to identify respondents for the main data collection survey.

---

[1] Most of these training materials we for internal training at our company, Lone Star Analysis. In addition, we sought high quality and referenceable information for clients in several government agencies and corporations who were to be trained

[2] We used APQC benchmarking processes, and the APQC benchmarking ethical standards

[3] The IEEE, NDIA (I/ITSEC), SPE, and Probability Management all allowed us, or actively supported data collection.

[4] The survey instruments were promoted to a wide range of organizations; industry, government, academia via the organizations who promoted the benchmarking and allowed use of their online communities.

Our second line of inquiry was interviews. We conducted interviews using a tiered approach. We conducted initial screening to determine whether an individual would be able to speak knowledgably about MS&A practices in their organization, and whether the organization had any meaningful potential to be among the best practitioners. A "pre-interview" came next. It targeted a few areas of common weaknesses. Here too, the aim was to screen out organizations with little chance of being best practitioners. A summary of these candidates is shown in Figure 1.

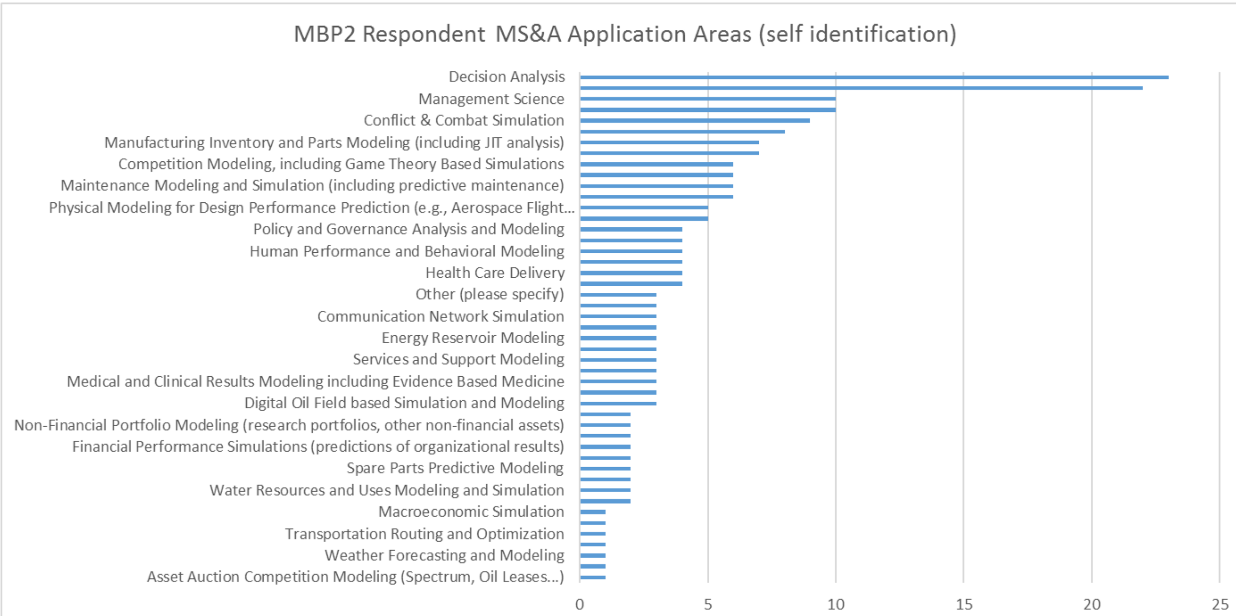

Figure 2 –Survey Respondent Self-Identified Application Areas

Based on a literature search (see the References section below) and benchmarking, ten topics were chosen[5] to provide rich benchmarking discussions and surveys. Identifying best practices depended on creating this candidate list of practices, and attempting to understand how best practitioners deployed them:

1. MS&A Design Environment for Low Cost, High Value
2. MS&A Capacity to Provide Transparency (Glass Box Models)
3. MS&A Capacity to Support Open Interfaces
4. MS&A Capacity to Accommodate Complexity
5. MS&A Capacity to Accommodate Diversity
6. MS&A Capacity to Accommodate Uncertainty in Cognition, Representation, Computation
7. MS&A Capacity to Accommodate Audit & Validation
8. MS&A Capacity to Provide Security
9. MS&A Process Discipline
10. MS&A Capacity to Be People Driven; Subject Matter & Analysis Talent

For both interviews and surveys, respondents were provided with definitions of each topic area. Semantic discipline was critical to our work. It's not surprising that a practitioner who does petroleum reservoir simulations uses different semantics than an aerospace structures engineer. The diversity of respondents to surveys is shown in Figure 2.

In addition to questions about the processes and disciplines used by our respondents, we also explored the types of MS&A they were doing. Our respondents were diverse, representing more than 40 different disciplines and applications of MS&A, and respondents came from several nations. For interviews, 63 potential exemplary

[5] We saw lists as short as five topics, and others with scores of topics. We created both the first list of ten, and the final list of fourteen using compactness as the key criterion.

organizations were identified, and over 100 individuals were approached for multi-stage interviews. For surveys, we obtained the cooperation of professional societies with memberships of more than 200,000 people. Using social media communications, we believe we targeted about 20,000 qualified respondents. The survey was lengthy and respondents knew we sought best practice[6] (not typical practice). We expected a low survey response rate, and in the end, we obtained forty usable responses.

Of these respondents to initial interviews and surveys, twelve were invited to the final stage of interview and four best practice exemplars were identified. Two agreed to be acknowledged: The U.S. Energy Information Agency (EIA) and the U.K. Metrology Office.

THE BEST PRACTICES

As the work progressed, we reported our progress at technical conferences and symposia[7]. We invited critique of the initial list of best practices for completeness. We developed a list[8] of reading materials helpful in assessing best practices. We reviewed our interview findings with attention to topics which seemed important to the respondents, but which were not part of our formal topic list. As a result, the list grew to the fourteen topics which follow.[9]

1. ***Intended use** – Is it clear why modeling is being done, and when it needs to occur? Does the using organization have clear policies regarding when modeling and analytics are required, and standards which apply to different classes of analysis? Is mission critical and safety related modeling held to a standard aligned with risks?*

During our work, we met individuals whose work was critical to safety, financial soundness, and national defense. But, only a few high-performance organizations exhibited clarity in this area of “intended use.”

We found a range of practice in intended use. We found many organizations use the methods and tools which are “standard” even when they do not apply.[10] Barriers to high performance we observed were:

- Starting late: best practice exemplars had clear time-lines on “when” results were needed and “when” MS&A should begin. Lower performance organizations often started late. Even organizations with policies requiring analysis often failed to enforce their own standards because they started late.
- Standards confusion: lower performance organizations exhibited confusion about why certain tools and methods were used.
- Use case confusion: lower performance organizations struggled to explain why analysis was being done, and how decision makers would use analytic results.

Best practitioners could explain in detail how often they needed to run their models, who the MS&A was for, and why the end users needed it. Use case clarity seems to be a critical feature of best practice.

---

[6] The APQC benchmarking guidelines make a distinction between median or typical practices, and “best practices” and this provided the initial framework. As the study progressed, we used in-process open peer reviews to refine what was deemed “best practice”

[7] Probability Management was most supportive, promoting discussions and presentations in 2016, 2017 and 2018 annual meetings. In addition, that organization created a Best Practices Chair and Program. IEEE and SPE provided their web forums for best practices discussion. The SPE is a joint sponsor of the PNEC conference, where an in-process paper was presented in 2016.

[8] The list is presented in the references at the end this paper.

[9] In early 2017, shortly after the last interview, interim results were presented at a Probability Management conference. The attendees were sophisticated analysts who critiqued the ten-topic list, and helped finalize the fourteen topics used. Obviously, there are other ways to organize our findings. However, this list has the advantage of being the most compact construct which seemed complete.

[10] As in other findings, we are not alone in this observation. Clarity of purpose is foundation of Ronald Howard’s Decision Analysis work at Stanford. J. Scott Armstrong’s 2001 paper, *Standards and Practices for Forecasting* begins with this principle (out of 139 practices). Retrieved from http://repository.upenn.edu/marketing_papers/135

2. ***Semantic Clarity** – Is there agreement on what words mean*[11] *and which measures are preferred? Do users and decision makers have clarity about which factors are direct, objective measures, and which are computed, or estimates? Is it clear what the model is, what inputs are, and what output data is?*

Survey respondents said there were an average of six distinct academic disciplines represented in their typical MS&A effort. A discipline might be finance, psychology, physics, aerodynamics, geology, or medicine. These disciplines use different semantics and measures. It is possible our respondents, interested in best practices are doing more complex MS&A than most other analysis professionals.

But less than 20% of respondents say they are highly flexible in semantics and measures.

Most respondents who reported large models, with multiple disciplines also reported poor flexibility in accommodating semantic needs of their stakeholders. It seems the analytics in greatest need of semantic clarity may be at greatest risk. Despite this, a minority reported semantic flexibility (Figure 3).

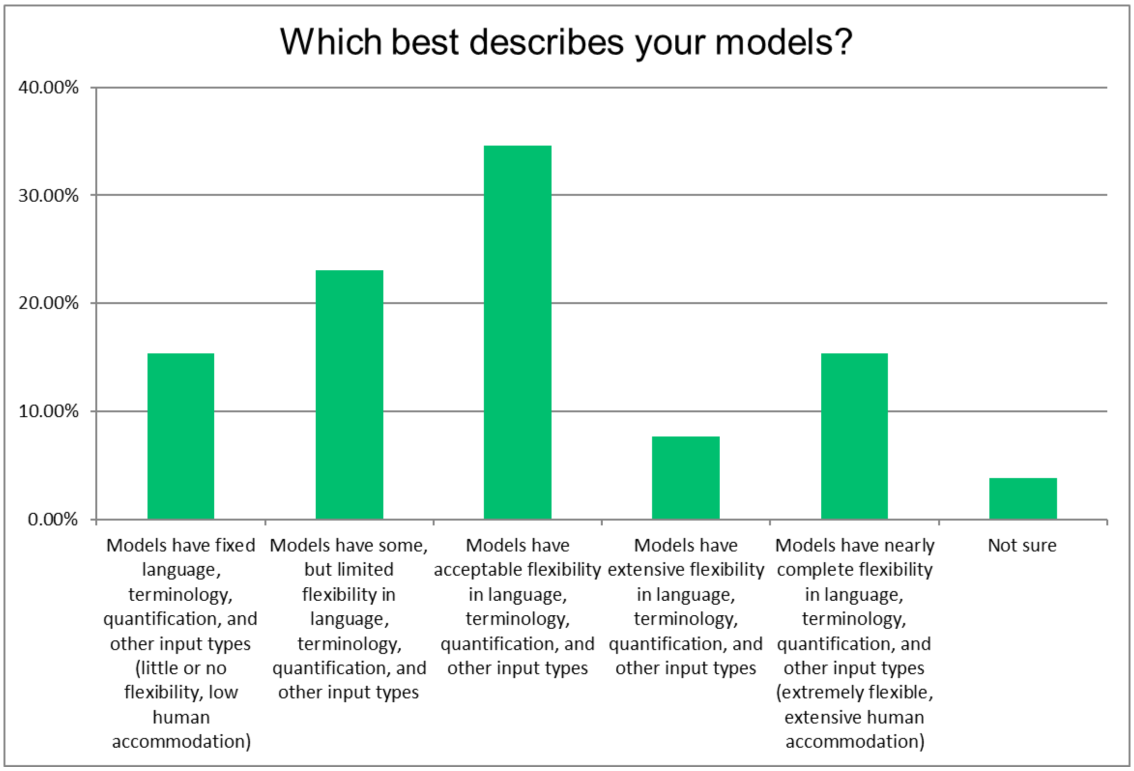

Figure 3 – Survey Respondent Semantic Flexibility Usually has Limits

High performance organizations took pains to communicate using semantics familiar to their target audiences and stakeholders. All four best practitioners spoke about this.

3. ***Design Environment for low cost, high value** – Do tools enable rapid and cost-efficient development of models?*

Respondents at all performance levels thought modeling platforms and tools were very important. Interviews showed best practitioners thought a great deal about how to use tools effectively. Best practitioners could explain their limits and challenges. Best practitioners were attacking these barriers to make their work better at the same or less cost, or to simply lower costs. All best practitioners involved executive leadership of their organizations in this planning. Three of the four could describe how long-range budgeting and planning connected technology budgets and organizational goals.

[11] Even the term "modeling" needed definition. We told interview subjects that we defined "modeling" as "computer abstractions of reality" including "simulation" or "forecasting" and "big data."

Low performers could only explain what they did and what they liked about their tools.

4. ***Process Discipline*** *– Are there clear model development processes which define how reviews are conducted, how configuration control is administered for both the computer simulation, and for the data sets?*

We found that less than 25% of survey respondents exhibited a high degree of process discipline. In fact, nearly 80% would be deemed to have no processes, by the standards of some quality surveys.[12]

In interviews, all four high performance organizations discussed processes and process discipline in detail. High performance organizations talked about peer reviews, user groups and active participation in their descriptions of process discipline.

Some low performance organizations also talked about processes, but there were important differences. Low performance groups seem to rely on either pseudo-processes, or, on red tape as a substitute for active, engaged, peer driven discipline. The problem of low quality, and lax process discipline seems to be widespread. A literature review confirmed others have reported similar findings.[13]

5. ***Transparency (Glass Box Models)*** *– Do those with a "right to transparency" have easy insight into how the model works, and what its limitations are? Is there clear understanding about who has a "right to transparency"?*

Best practitioners were concerned about transparency, low performers either didn't care, or relished being opaque. Only about 30% of survey respondents reported their models were open enough to potentially be "glass boxes." During the study, this issue was widely discussed in the press, and was the focus of regulators and EU law makers. The Association for Computing Machinery US Public Policy Council issued "Principles for Algorithmic Transparency and Accountability."

A Lone Star survey of US citizens shows growing sentiment for a "right to transparency." This is the right to know how models and algorithms work when they impact the well-being of ordinary people. EU GDPR took force in 2018, including a right to transparency about how algorithms affect EU citizens, mirroring sentiments shown in US polling.

Three of the best practitioners used some form of "signing your work," a phrase we used to describe attribution of analytics to a responsible person. They saw a connection between transparency, responsibility, and accountability. This connection was never observed in low performing organizations. Instead, low performance organizations tended to obfuscate by-name responsibility.

6. ***People Driven; Subject Matter & Analysis Talent*** *- Can real humans put data in? Do they "get" the answers coming out?*

This topic is at the intersection of several others; Open Interfaces, Accommodating Complexity, Accommodating Diversity, and Accommodating Uncertainty. However, reviewers felt this topic was distinct, even if related to other topics. Sooner or later a human will have to deal with the MS&A input, output or both. This is true even for IoT robotic systems, or real time buying/bidding/trading systems. But, it seems that less than 20% (probably less than 10%) of MS&A practitioners robustly help the humans (stakeholders) who are involved.

All the best practitioners talked about this, and making their products more meaningful, more usable, less confusing. Some of these organizations were involved in very complex analysis. It would have been easy for them to point to the sophistication of the topic, as a "reason" for poor user experience (UX).

In some cases, best practitioners struggled with making subtle and complex analytic results accessible to stakeholders who were often pressed for time. This is a difficult topic, and to some degree all the best practitioners expressed concern or self-criticism in this area. All the low performers seemed to ignore this, or even used bad User Interface (UI) as a moat to keep others out of their "castle."

---

[12] For example, the standards of "what is a process" as taught to Baldrige Award examiners, and to ISO-9000 examiners.

[13] Several of J. Scott Armstrong's publications survey the lack of discipline, and poor processes. A recent Harvard Business Review article's title is illustrative; *Only 3% of Companies' Data Meets Basic Quality Standards,* HBR, September 11, 2017.

7. ***Open interfaces*** *– Is it easy to get data in and out of the model without paying for special software, or paying for excessive processing, data cleaning, or other costly and time-consuming steps?*

Less than half the survey respondents claimed their MS&A inputs were compatible with common software titles and file types. About a third say their interfaces are not usable by others, or with common software titles and file types. This is a contrast to best practitioners. All four exemplars put significant emphasis on accommodating several methods for data entry and, for exporting results.

Survey respondents rated file compatibility as an important attribute of open interfaces (58% said this was the most important, or second most important attribute). Open interfaces were not an assurance of best practice. Some low performing organizations use open interfaces. Spreadsheet based analytics are an example. While spreadsheets are not bad per se, some of the most embarrassing errors we saw came from work in spreadsheets.[14]

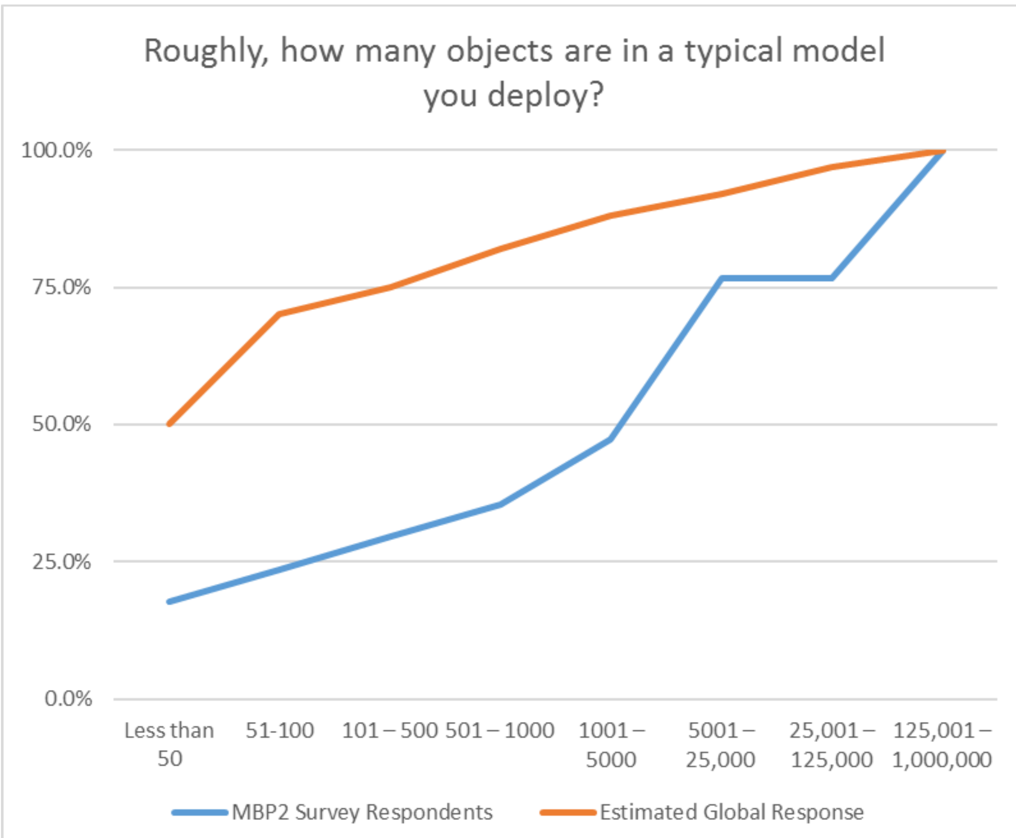

Figure 4 – Survey Respondents and Model Complexity

8. ***Accommodate Complexity*** *– Does the model adequately cope with real-world complexity and interconnections of systems represented in the model? Is representation overly constrained by limitations in the approach? Is complexity manageable, or is there what Box called "needless elaboration"?*

This is a complex and controversial topic. G.E.P Box warned against "needless elaboration." It seems clear that MS&A tools can be either too complex, or overly simple. The balance between "overly constrained" and "manageably complex" is subjective. Not surprisingly, about 77% of survey respondents judged their own efforts to be "about right" in this area, and none said their work was "far too simple" or "far too complex."

[14] See Carl Bailick's Wall Street Journal column of April 19, 2013; *Spreadsheet Slips Not Economists' Only Problem.* Bialick explains how Harvard economists used a flawed spreadsheet to generate flawed economic advice to several nations.

For benchmarking, we used the concept of a mathematical "object" to assess complexity.[15] It seems the survey is somewhat distorted by best practitioners who are competent to use larger tools. It seems likely smaller models are more common, and larger models are rarer than reported by our respondents, as shown in Figure 4.

The exemplars put great emphasis on this topic. One of them worked under highly constrained circumstances and had to devise innovative means to accommodate complexity with short deadlines and limited computing capacity. The other three exemplars had more powerful tools and fewer constraints, but still put emphasis on striking the right balance in complexity. Low performance organizations were guilty of making simple issues complex. Perhaps more commonly, interviews seemed to suggest they simplified complex problems to fit their tools and staff, even if the integrity of the analysis was compromised.

9. ***Accommodate Diversity** – Does the model accommodate different disciplines who may not use the same measures or the same semantics, such as finance and physics?*

This is related to topic #2, Semantic Clarity. But it is distinct. This issue deals with the analytic blending of models which originate in different disciplines. For example, physical failure analysis and prediction precedes a financial analysis of warranty claims. This topic was important to the best practitioners. The UK Met Office, for example integrates the work of many different types of science in their weather models.

Over half of the survey respondents report serious problems including limited choices in naming, "unnatural" choice of measures, units, and choosing one discipline over another (physics over chemistry, accounting over marketing…). This was not rated as an important topic among survey respondents. Only about 20% of survey respondents reported "nearly complete flexibility" or "extensive flexibility" in language, terminology, and quantification. This was a contrast between low performing organizations and best practitioners.

10. ***Accommodate Uncertainty (in cognition, representation, computation)** – Does the model incorporate the full span of mathematical uncertainty and is this preserved with correct computational methods? Is uncertainty represented in a manner consistent with the nature of the problem, and the nature of the available information? Is uncertainty provided to users in a way compatible with cognitive limits? Does the model do "the Arithmetic of Uncertainty" correctly?*

This topic was of great interest because nearly all analysis deals with some degree of uncertainty. Yet, about a third of survey respondents all but ignored uncertainty. This was an area where the benchmarking team found the most mathematical malpractice in literature reviews[16], surveys and interviews. It may be the area of greatest regulatory risk for some organizations. A key topic is choosing how to represent uncertainty as shown in Figure 5.

---

[15] The term "object" will be familiar to software developers and other users of object-oriented methods. This is an example of roughly 20 topics we had to carefully define for survey and interview purposes. Here, and "object" is either an input, or a process which results in a mathematical quantity, a Boolean state, or other analytic result, such as a classification.

[16] For example, see Siegfried, T. (2010). Odds are, it's wrong. Science fails to face the shortcomings of statistics. Science News 177 (7), 26-29. http://xcelab.net/rm/wp-content/uploads/2010/03/Odds-Are-Its-Wrong-Science-News.pdf

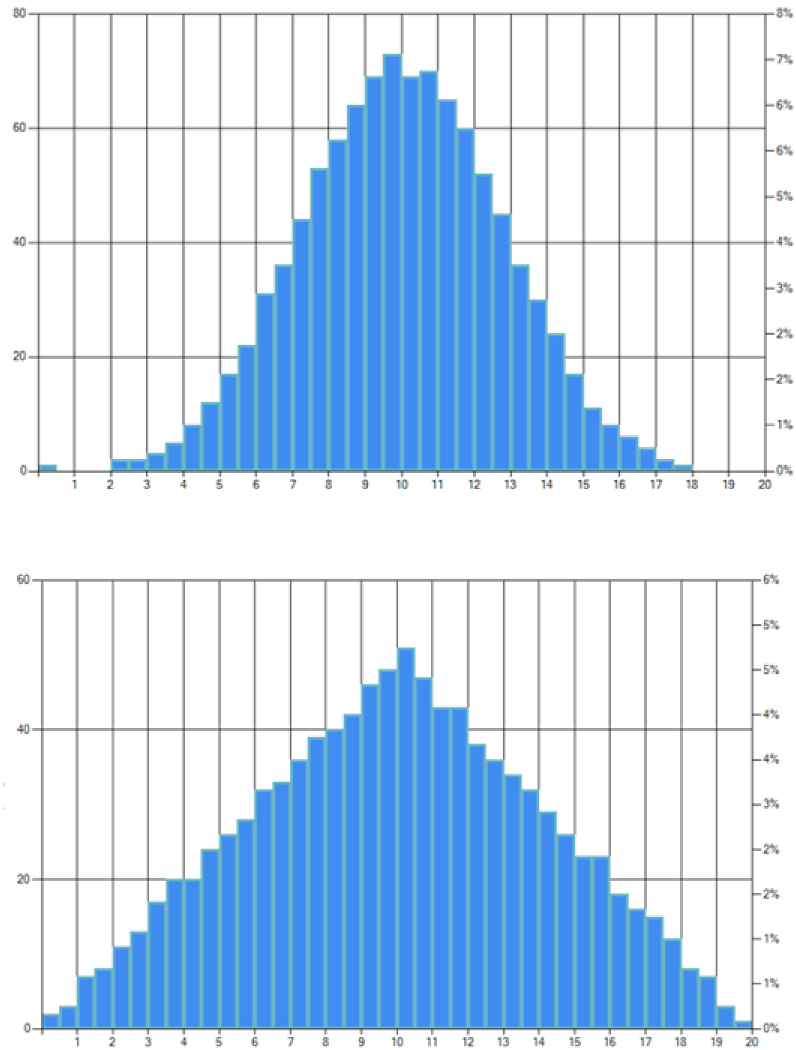

Figure 5 – Triangle Distributions are Not Gaussian

At least two government agencies we reviewed were guilty of "mathematical malpractice." Both claimed to be best practitioners in some sense. Both used single number proxies as a substitute for the full span of uncertainty they were assessing. In one case, we could audit the math of one of these agencies and find demonstrated significant errors. The agency was provided feedback but seemed to ignore it. In another case, a group we call "the highly regarded analytics organization" (THRAO) said they used tringle distributions, because those were 'good enough' for their purposes.[17]

But, how uncertainty is represented is important because it changes the MS&A results. For example, consider two distributions, both with ranges from zero to 20, and both centered at 10. Both have a mean, and median value of 10. These are shown below.

What happens if we multiply a 2, 4 or dozen of each distribution? We know what happens when we multiply the "expected value" of ten:

10 x 10 = 100

10 x 10 x 10 x 10 = 10,000

[17] One of the government agencies appeared to be violating both OMB directives, and Title X statutes. THRAO seemed to violate SEC reporting guidelines.

10 times itself 12 times is 10^{12} or, one Trillion (1T)

Will we get answers like that if we multiply the distributions instead of the single number representation? Will the Triangle and Gaussian be pretty much the same when we do these multiplications?

No. None of these are the same. The upper figure is a set of Monte Carlo Trials from a Gaussian distribution, and the lower figure is a Triangle distribution. Both have an expected value of 10.

The median value of the distribution which results from multiplying the twelve Gaussians is 6.3×10^{11}. In contrast, the median value of the distribution which results from multiplication of twelve Triangle functions is less than half the same result for Gaussians; 2.96×10^{11}.

So, THRAO risks reporting values which are quite different than the best estimate. And, government agencies using single number proxies are also reporting results far from the best estimate.

20% of survey respondents said they "preserve the span of uncertainty across all objects in the model without using single number proxies." So, these failings are common. All but one exemplar preserved full spans in most cases. The EIA was unable to do this but used mechanisms to cope with the issue.

11. ***Accommodate Audit & Validation*** *– Does the process ensure error detection and correction is done? Do the modeling processes and tools make it easy to conduct audits, and to archive the results? Is there a consistent commitment to this discipline?*

This area was perhaps the most shocking of our benchmarking. Only one in six respondents always conducted audits before publishing findings. Auditing was a statistical tie for the least important topic in our survey. Apparently, MS&A practitioners avoid checking their work, or see this as unimportant. This was consistent with anecdotal findings.[18] Over 40% of survey respondents said they never or rarely did audits unless the customer paid. Over 80% fell short of a "disciplined process" for checking results. We make a distinction between checking MS&A results, and validating a simulation tool. "Validation" was an important issue in some applications. But, we found misunderstandings about what validation meant.

All four exemplars demonstrated significant discipline in this area. Forecasts and projections are often impossible to fully audit. In those cases, the MS&A organization has the responsibility to take steps to deal with the lack of testing. The exemplars all exhibited some combination of these attributes:

- Test what can be tested
- Offer an estimate of uncertainty about the forecast or projection
- Check accuracy when forecasted events become history
- Attempt to advise consumers of analytic results with insights to analytic risks

A disciplined distinction between the extent to which a tool is valid, and believing a specific set of results are correct

12. ***Provide Security*** *– Does the system (the model, the people, the computing equipment) provide security and privacy protection adequate to comply with applicable obligations, and to protect stakeholders?*

Security seemed the most widespread best practice. Exemplars could speak in detail about what they did, and why. But, a typical survey respondent reported an average of three separate security practices. Only about 15% of respondents reported no security measures. Because security varies with the type of MS&A, and with data, it is difficult to make broad observations. In some cases, security is defined by what is not allowed. Privacy and security can mean some data is not collected, or not used. Software "white listing" can restrict use of some software titles. For example, R and Python are often not usable in high security defense environments. Network limitations can also be the result of security measures.

13. ***Processing and Network Compatibility*** *– Do processing loads and data flows associated with the modeling fit within the time and cost constraints of the modeling purposes?*

[18] An "expert" said, "When I price out the cost of auditing results, customers never want to pay for it, so I don't bother doing it."

High performance organizations thought a great deal about this topic. Low performance organizations seemed to assume computing power and network capacity would just be ok, or was "someone else's" responsibility. Exemplars could describe their approach to this topic in some detail. The Met Office owns several super computers. Two exemplars operated with highly constrained computing resources. It seemed significant that the best practitioners felt this was a critical topic, even when their computational resources were so different. A near-best practitioner also spoke about this topic as a risk in terms of cloud computing. They warned against magical thinking that the cloud would somehow make all processing power problems evaporate.

Low performing organizations seemed oblivious to these problems. The "curse of dimensionality" was often ignored. In our benchmarking, it seemed "Big Data" and AI advocates tended to underplay the difficulty of scaling. Academics and consultants seem to be less sensitive to this topic than practitioners who must cope with it.

14. ***Statutory and Regulatory Compliance** – Are obligations clearly understood and is compliance documented? Are all the model stakeholders aware of obligations which might be associated with the model and its results?*

We presented a list of more than a dozen regulatory and standards bodies who are the basis of law and regulation for SM&A. The regulatory area we tested was mathematical treatment of uncertainty. This applies to a range of regulated behavior in government. and business. We asked, "Is your modeling subject to the guidance or requirements of these, or similar organizations, regarding the representation of uncertainty?"[19] Over 80% of respondents said "No" or "Don't Know." Less than 18% said "Yes." Based on the self-description of the respondents, we expected more than 80% should have said, "Yes." If that assessment is correct, then most of our respondents seemed guilty of regulatory problems. Oddly, our pre-interview work showed government agencies are among the least compliant. It also seemed that agencies with regulatory authority were the worst offenders in terms of regulatory violations[20].

During the project we saw a growing discussion of transparency, and the EU's GDPR regulations.[21] Best practitioners and near-best were very sensitive to these matters. Self-assessments in this area seem to be the lease reliable of the 14 topics. Third party assistance may be needed to evaluate regulatory and legal compliance.

SUMMARY/RECOMMENDATIONS

Our work shows clear contrast between high performance practitioners of modeling, simulation, and analysis compared with most others. We believe the fourteen topics presented here can provide a template to improve nearly any MS&A effort.

As part of our literature review, we identified a few resources useful for those who want to improve their performance. The reader will find them in the References section below. We recommend these as a starting point for organizations seeking to improve their performance. We also recommend using the 14 topics presented here for self-assessments.

We are working to finalize our risk checklists and plan to publish additional findings to make the benchmarking more easily accessible.

As the importance of algorithmic results grow, and society is increasingly dependent on them, we believe practitioners have a responsibility to improve. We hope our work is useful to them.

REFERENCES AND RECOMMENDED READING

---

[19] The full question was, "A number of organizations issue guidelines or specifications for the representation of uncertainty. They include Society of Petroleum Engineers, U.S. Office of Management and Budget (OMB), European Medicines Agency (EMA), U.S. FDA Office of Regulatory Affairs, U.S. Federal Reserve, U.S. Office of the Comptroller of Currency, Bureau International des Poids et Mesure (BIPM), International Electrotechnical Commission (IEC), International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), International Organization for Standardization (ISO), International Union of Pure and Applied Physics (IUPAP) and, International Organization of Legal Metrology (OIML).Is your modeling subject to guidance or requirements of these, or similar organizations, regarding the representation of uncertainty?

[20] See Appendix IV

[21] Topic 5, above

Armstrong, J.S., et al. (2015). *The Golden Rule of Forecasting: Be Conservative*. Journal of Business Research. (The article proposes a "unifying theory" of forecasting with a checklist. Professor Armstrong has published a body of work at the Wharton School. Many of his papers are available at the University website.)

Barth. R., Meyer, M., and Spitzner, J. (2012). *Typical Pitfalls of Simulation Modeling*. Journal of Artificial Societies and Social Simulation. (This paper was an early influence for our work, describing five logical reasons why modeling, simulation and analysis fail to deliver good results.)

Hubbard, D. (2010). *How to Measure Anything: Finding the Value of Intangibles in Business*. John Wiley & Sons, Inc.

Kahneman, D. (2013). *Thinking, Fast and Slow*. Farrar, Straus and Giroux. (Perhaps the most important and readable book on human biases.)

Law, A.M. (2009). *How to Build Valid and Credible Simulation Models*. IEEE Winter Simulation Conference. (This is a rare paper which attempts to think about modeling, simulation and analysis generically.)

Savage, S. (2009). *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. John Wiley & Sons, Inc. (One of the most readable books on quantitative analysis.)

Wasserstein, R., and Lazar, N. (2016). *The ASA's Statement on p-Values: Context, Process, and Purpose*. The American Statistician Vol. 70, Issue 2. (Available on ASA's website.)