

Risk Assessments for the Age of Algorithms

Steven Roemerma John Volpi Randall Allen, PhD

Lone Star Analysis
UNITED STATES OF AMERICA

sroermerman@lone-star.com; jvolpi@lone-star.com; rallen@lone-star.com

ABSTRACT

The authors led a benchmarking effort crossing disciplines: decision analysis, operations research, risk modelling, management science, conflict and combat simulation, and logistics and supply chain simulation. Practitioners volunteered to describe their practices and learn from others. While different practitioner communities have agreed on local practices, there is little interaction among communities.

Previous publications described best practices highlighted from benchmarking. Two gaps were discovered: a disturbingly high fraction of poor practice and the absence of executive-level risk assessments. Executives often lack the time and/or technical background to form risk assessments for analytics provided them.

This paper offers a new and simple means for executives to assess risks. Six non-technical questions address most of the risks seen in the benchmarking. The method is based on a checklist founded on the international benchmarking effort. The authors also conducted research into specific risks, including legal risks which arise from increased reliance on analytics. Among the risks are some unique concerns related to Artificial Intelligence.

The work to identify risks showed hazards arise from several sources and produced a risk checklist which does not require in-depth Modelling, Simulation, and Analysis (MS&A) knowledge. This paper presents the checklist, as well as some of the deeper MS&A principles which support it. This is useful for both executives and practitioners.

The Research was supported by several professional societies, industry groups, and non-profit educational associations, including the Interservice/Industry Training, Simulation, and Education Conference (IITSEC), the Society of Petroleum Engineers, the Institute of Electrical and Electronics Engineers, and Probability Management.

1.0 SUMMARY OF THE STUDY AND PREVIOUSLY PUBLISHED FINDINGS

We have previously published (Roemerma et al.) the nature of the study and our data collection. To recap, in 2014 and 2015, the authors proposed a cross-domain benchmarking study. As practitioners working across several domains, we noticed “normal” simulation and modeling practices in some fields were unknown in others. We proposed a multi-domain study to several organizations. Universally, the feedback was positive, but none wanted to lead the effort.

Eventually, we decided to proceed on our own and began to recruit help. We were aided by many of the organizations we had approached, and others joined along the way.

Risk Assessments for the Age of Algorithms

- The Institute of Electrical and Electronics Engineers (IEEE) was the first to allow us to use one of their web forums to discuss the study and solicit participants.
- The Society of Petroleum Engineers (SPE) granted us the same web forum access.
- The Wharton School of the University of Pennsylvania saw two Professors support the study with advice and access to unpublished literature and help from graduate students.
- INFORMS provided access to their online forum.
- I/ITSEC allowed us to survey portions of their membership.
- Probability Management gave us access to their membership and then named one of us Chair of Best Practices.
- The corporations Lockheed Martin and Chevron provided us access to their modeling and simulation communities for interviews and surveys.

In all, these organizations had memberships of about 200,000 individuals (not including the two corporations, whose employees were presumably members of the societies we approached). Of these, we estimated less than 10% of the members were active practitioners in Modeling, Simulation, and Analysis (MS&A). Of those, we estimated that about 2,100 of those saw our invitations to participate in the survey and interviews.

Beyond working with these large groups, we initially targeted more than 40 individuals for participation in our data collection because of the reputation of their organizations or their personal reputation for excellence. We solicited these by-name targets and they proved to be a rich source of information. In the end, we approached 126 individuals from 65 organizations across many domains (see Figure 1).

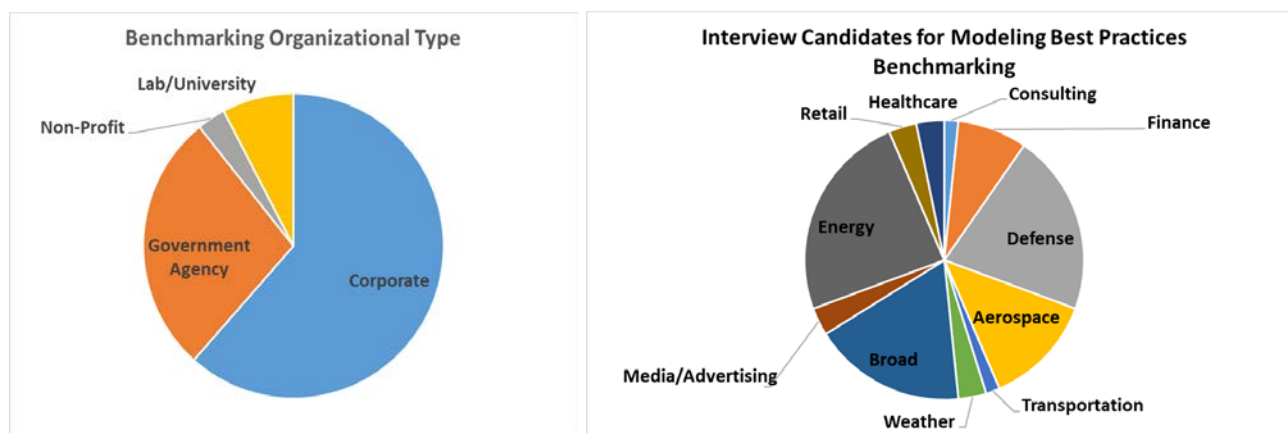


Figure 1: Participants Were Drawn from A Variety of Domains

We used two principal survey instruments and an interview instrument. Both the principal survey and interview were long and required a significant effort from participants. In the end, we were surprised to see more data collected from the interviews (126) than useful survey responses (40). Overall, our survey response rate was about 2%, respectable for a long, tedious survey. Because some individuals were both interviewed and responded to the survey, we estimate the total number of individuals who participated at about 150. These responses were international:

By 2016, we gathered enough preliminary information to publish some early findings and to present at conferences.

In 2017, we presented a preliminary summary of our findings at the Probability Management Annual Conference in San Jose. A special session of the conference was used to peer review the best way to list and explain the best practices. The result was fourteen best practices, grouped in three categorical themes:

- 1) Strategy & Architectural Approach
- 2) Implementation Disciplines
- 3) Execution Disciplines

At the same conference we described four organizations who best exemplified and utilized the best practices: The Energy Information Agency (EIA) of the United States, the Metrology office of the United Kingdom, and two organizations who chose to remain unnamed. We dubbed them “A” and “Z.”

1.1 Key Findings Previously Published

The most comprehensive peer-reviewed publication of the best practices work appeared at I/TSEC 2018 as a paper and presentation. All fourteen practices and how they fit into three categories were discussed. Our findings showed that across most of the fourteen best practices topics, less than half of MS&A practitioners were exercising any single best practice. Being consistent in all areas is clearly exceedingly rare ($50\%^{14} = 0.006\%$).

This was a discouraging result because interview targets were chosen because someone claimed they were good to excellent, and survey respondents are biased toward better practitioners. We concluded the typical state of the MS&A practice is worse than we saw in our sample. This poor showing led to the exploration of risks. In summary, we concluded:

- Risks were poorly understood by practitioners and by executive users of MS&A.
- Risks were not merely the absence of best practices.
- Artificial Intelligence as a special class of algorithms presented special classes of risk.

Few areas of business and government can operate without some form of MS&A today. Complex topics require analytics to support decisions. Automatic algorithms run everything from securities trading to suggesting maintenance actions. Clinical trials, oil exploration, aerospace design, and flight training all depend on simulations. Yet, the executive user of MS&A results faces a serious challenge. While the executive must rely on MS&A to be competitive, few managers have the time or education to assess the risks posed by the MS&A used by their organizations.

We noted differences in the way high and low performing MS&A groups described their processes. As shown in Figure 2, even when good practices are used, low performers use them differently than high performers. This led us to create a risk checklist for executives, based on topics consistent with reasonable executive oversight. A full description of the risk checklist and its origin has not been published before. This paper addresses that gap.

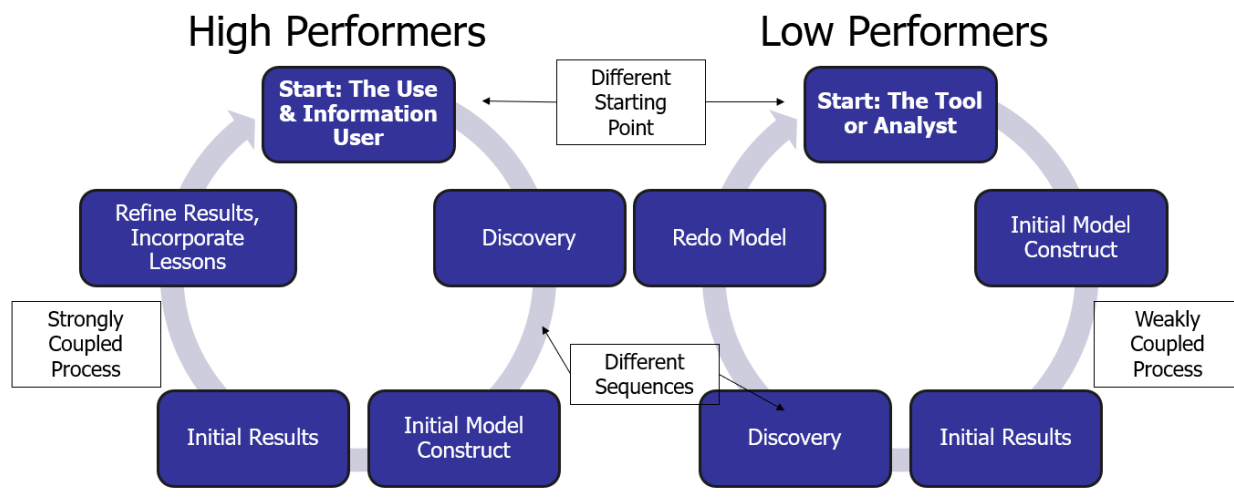


Figure 2: Some Differences Between High and Low Performance

1.2 Risk Identification Process

Early in our benchmarking process, the need to consider risks became evident.

A pair of questions in our survey instrument and in our interviews asked about the legal requirements related to MS&A. We had done research to understand the laws and regulations related to different industries and technical disciplines. In particular, we noted that dealing with uncertainty was specified in many different disciplines. We asked the following question:

A number of organizations issue guidelines or specifications for the representation of uncertainty. They include the 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?

As this legal risk became apparent, we added additional questions in our interviews. Only 12% of interview respondents were found to be free from legal or regulatory violations. So, both survey findings and interview findings indicated substantial organizational risks, which executives might not understand. We found that in some cases, the MS&A practitioners knew they were violating a regulation with serious legal consequences. However, they offered several “reasons” why it was “ok.”

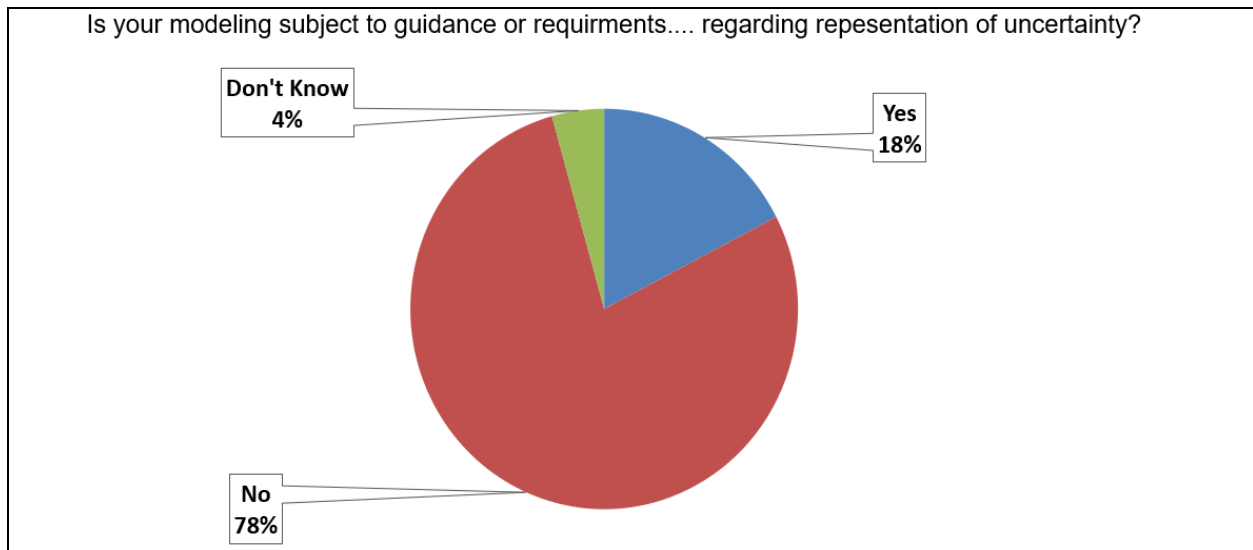


Figure 3: Most Respondents **NOT** Subject to Regulation in Treatment of Uncertainty

One anecdote helps to explain what we found in one of the industries we benchmarked. In the Petroleum industry, upstream producers must estimate reserves of oil in the ground, which are assets on their balance sheet. This is done using sophisticated reservoir modeling software. The Securities and Exchange Commission (SEC) and the Financial Accounting Standards Board (FASB) require these companies to state the reserves based on certain probabilistic standards and rely on industry professional society guidelines. In addition to the official guidelines, societies like SPE publish training materials, with a key issue being the shape of the uncertainty (the type of probability distribution).

A software vendor with a large market share in reservoir modeling agreed to an interview. We asked which probability distribution was favoured. It was easily chosen with a drop-down menu. The head trainer for the organization told us that the software defaulted to a distribution NOT approved by the SPE (but easy to use) and that they had never seen a customer switch to a “correct” distribution. Interviews with oil company users seemed to confirm this observation.

While legal risks first alerted us to the need for executives to be wary about hazards lurking in the MS&A they consumed, other discoveries further highlighted risk as a topic worthy of separate consideration.

Our agreements with those we surveyed and interviewed preclude us from naming them, but one low performing organization we can cite is the International Energy Agency (the IEA). Shown in Figure 4 are 12 years of IEA forecasts (coloured lines) and history (bold black) of Photovoltaic Energy Additions. In other words, the number of Gigawatts of PV power added worldwide each year. Clearly, the IEA is consistently using a flawed forecasting method.

Energy market forecasting is hard. No organization can hope to be highly accurate. But IEA problems are so persistent that it raised questions about the contrast between the IEA and the EIA, one of our best practitioners.

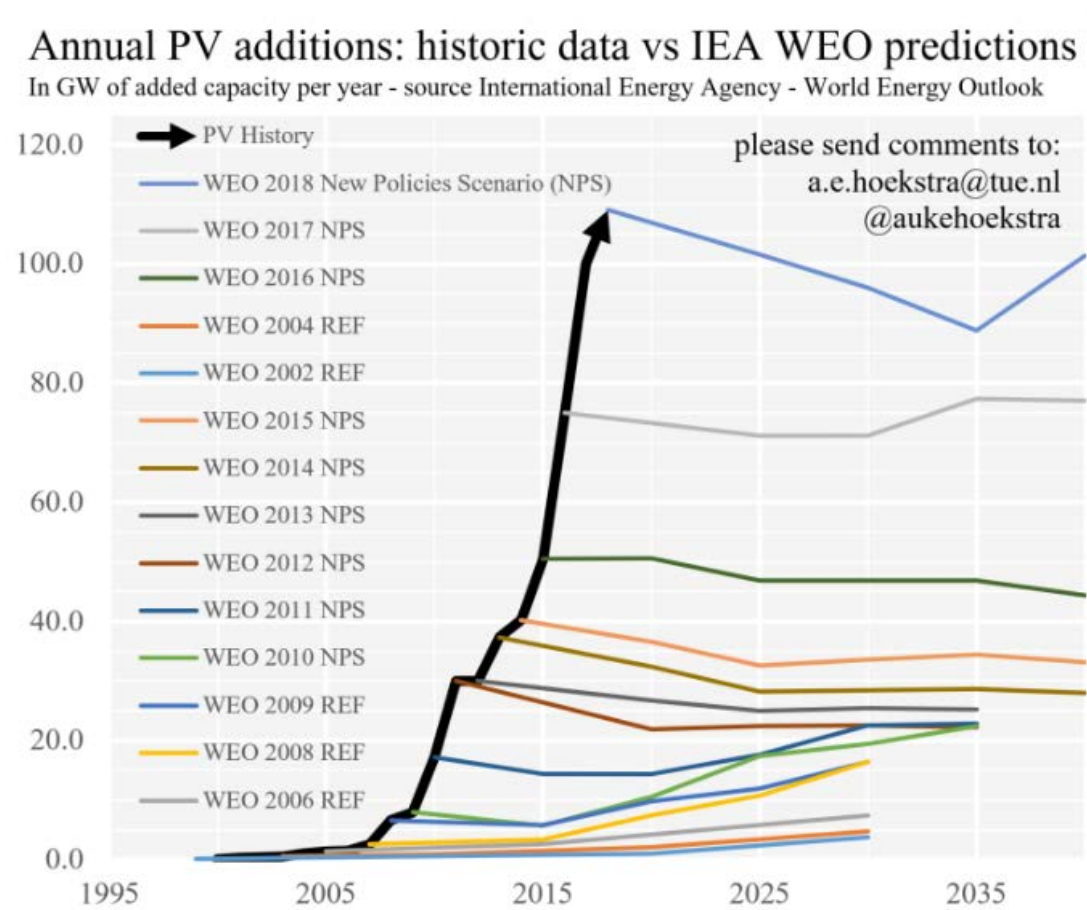


Figure 4: Some MS&A Groups are Consistently Inaccurate.

The contrast of the EIA and IEA was part of another trend we noticed. While three of the four best practitioners were government agencies, all the very worst practitioners were also from the government. Some were so bad that we dubbed their self-described methods as “mathematical malpractice.”

Because nearly 30% of our respondents were government agencies, we used this cadre to explore the concept of risk identifiers, separate from best practices or their absence. In particular, contrasting the 17 government groups we interviewed provided us with richer insights than broadly assessing survey responses alone.

2.0 EXECUTIVE RISK CHECKLIST

We compiled a list of proposed hypotheses for risk indicators and risk mitigators. We tested the hypotheses against the data from respondents. In many cases our small data set meant our correlation statistics had marginal value. As a result, many hypotheses were untested. What follows is the list of findings we felt to have value and validity.

The following six attributes are arranged from least diagnostic, ending with most suggestive. We used three methods to correlate each hypothetical risk indicator; quality of MS&A results, degree of process discipline, and a weighted average of the two biased toward results.

2.1 Repetition

One theory was that repetition leads to superior performance. But we saw no correlation between “one-of” MS&A work products vs. repetitive outputs and best practices. Among our best practitioners in government, we found some organizations doing repeated forecasting and others doing unpredictably eclectic MS&A.

Among our worst practitioners, we found roughly the same diversity. This was something of a surprise. The correlations were 26% to -39%, and this split is one reason repetition alone is a poor predictor. It was clear many respondents did not use repetition to improve their processes and results. We thought repetition might lead to more process discipline, but the emphasis on practitioner excellence seems unconnected or very weakly correlated.

Repetition alone is not a risk mitigator. Claiming repetition alone to be a risk reducer is misguided at best. Repetition may reduce risk if some third party (or the executive user) keeps score on the accuracy of predictions.

2.2 Budget

This theory suggested that better-funded MS&A teams would perform better because they had adequate resources. However, we saw only a weak correlation between the budget and good practice. Agencies with larger budgets for modeling, simulation, and analysis could not be proven to be “better.” In most cases, it was impossible to determine budgets, so the assessment was qualitative at best. MS&A spending is rarely broken out separately.

One of the unnamed best practitioners operated with very limited resources. So, if there is a connection between resources and process discipline, it is weak. One high-performance organization had high standards because it was resource-limited. They seemed to feel process discipline to overcome budget challenges.

Budget was the weakest of the suggested risk indicators with adequate funding being correlated 15% with results. But oddly, budget was negatively correlated with process discipline (-32%). So, here too, we saw mixed results in predictive power. At any funding level, the MS&A providers have limitations. What seems more important than budget per se was whether the practitioners could explain whether limited funding, time and staffing were constraints. Severe underfunding (or lacking the resources of time and talent) is a risk indicator, but even modestly funded MS&A teams can do excellent work.

These first two hypotheses are worth considering by executives but lack predictive power for quick assessments. We list them because they were frequent questions, unlike other hypotheses we rejected or found to be untestable. The following four had more value.

2.3 MS&A Mission

The hypothesis here was that a team dedicated to a specific MS&A mission was better than a team whose role included other responsibilities (not pure MS&A). As specialists, we have some bias toward belief (or at least hope) in this theory. We saw a weak correlation between a mission focused on MS&A and best practice. Across our benchmarking, all our best practitioners were MS&A specialists. We also noticed the best MS&A specialists could advise their leadership on which efforts should be outsourced and which should be retained.

Within the cadre of government agencies, there was a striking finding. The degree to which an organization had a team with a dedicated MS&A mission was negatively correlated with process discipline (-74%). For example, the worst offender was part of an organization whose only mission was to provide analysis. This “worst

offender” was described in pre-interview notes as “Guilty of mathematical malpractice, and proud of it.” So, specializing in MS&A is no guarantee of quality. Instead, it can become a license for poor practice.

This worst offender resisted all attempts at meaningful outsourcing, which suggests a diagnostic question, “has this MS&A organization advocated for bringing in meaningful outside support?”

2.4 Use of Static Tools and Methods

We saw a correlation between frozen MS&A tools and methods and risky poor practice. The degree to which frozen tools and methods were used as the third best predictor of bad results (29%), was weakly correlated poor processes (18%), and for the weighted process/results metric (-32%). Static tools and methods themselves may not be the root cause. They were often connected to monopoly providers, lacking accountability. Hence, this seemed to be a weak proxy for accountability. But frozen tools and methods also resisted process improvements.

The worst offenders tended to use tools which were both “validated” and “standard.” However, the tendency to want “certified” MS&A seemed to be the most dangerous of these risk factors. Here is a list of the poor practices we found related to static methods and tools.

- **Overreliance on “p-values,” which are supposedly a measure of “significance.”** We found widespread misuse and misunderstanding of procedures and “canned” software to compute p-values, and ignorance (perhaps willful ignorance) of the growing revolt against the use (and misuse) of this measure. For example, a spokesperson for the American Statistical Association said, “The p-value was never intended to be a substitute for scientific reasoning.” Yet, it remains easy to put data into a standard tool to generate a p-value, and then assume it means something, or worse, assume it means everything. P-values often mean next to nothing.
- **Overreliance on “our standard software.”** This was related to reliance on Validation, Verification, and Accreditation (VV&A) (discussed below) but was a broader observation. Organizations with specialized tools tended to rely on the tools, even when they were ill-suited for some applications. In other cases, cybersecurity practices made it difficult to select the best tool for the job. No one disputed the benefits of established standards, tools, and methods. However, being trapped in a decades old toolset is not “standardization” in most cases. It is stagnation.
- **Overreliance on “our approach.”** This tended to be a combination of process and organizational stagnation. Organizations relied on the same people and the same processes, even when they were ill-suited for some applications. These organizations were not adaptive even when their stakeholders faced dynamic challenges.
- **Overreliance on “our single approach and expert.”** This is the extreme form of the prior two problems. Three organizations we considered all exhibited a very dangerous pattern; an old Microsoft Excel spreadsheet which had become the de facto official tool, and one person who was using it without ongoing peer review, audits, or testing. In all three cases, others in the organization knew the spreadsheet and the analyst were flawed. But these organizations had become dependent on the static tools and felt they had no real alternatives.
- **Overreliance on formal validation.** We saw what seemed to be a significant connection between formal validation and poor practice. We spoke with instructors at prestigious government universities. One was the first to suggest reliance on formal validation was, at best, a placebo, and at worst, permission for malpractice. Broadly, we found this to be confusion about the audit of a tool and the audit of results. Good results are not assured because of some past assessment of software.

No expert we spoke with felt formal “Validation, Verification and Accreditation” was as useful as claimed. This was true for MS&A, which fell under U.S. DoD Instruction 5000.61 and other forms of “official” validation. No one disputed the benefits of audits and peer reviews. But lengthy VV&A seemed to be a poor value. One senior government executive we spoke with said her organization was required to use tools with VV&A, even though they were several years old and no longer represented the reality of her department. Billions of dollars were being allocated based on these tools. She was frustrated because policy required the use of VV&A, which in turn meant three years to change toolsets. It was simply too hard to change.

More disturbing, a pre-interview with one analyst in a government agency exposed the following. This individual claimed ONLY they could perform certain assessments because they had the ONLY approved tool, which created a monopoly and personal power for this individual.

2.5 Misuse of Averages and Single Number Proxies

We saw what seemed to be a clear pattern which related the use of single numbers with poor practice. This was tied with accountability as the best predictor of risk we tested; 92% correlated with bad results. It is related to a benchmarking topic, “Accommodate Uncertainty” (in cognition, representation, computation). It was not correlated with poor process discipline. It makes sense – a highly disciplined process is distinct from the quality of the process. Cargo Cults showed extremely good process discipline as they asked the gods to send them loot from the sky.

But beyond bad math, we suspected we observed apparent willful misrepresentation by the use of averages. For example, an agency we benchmarked published a proposed rule in the U.S. Federal Register seeking public comment (as required by law) and “estimated” the average time to comply with the new rule (also required). The agency estimated ‘the average firm’ would need 15 minutes to file the necessary reports. It appears this “average” is sleight of hand. The “average” firm is one which is 80% exceedingly small (who do not file and have zero time to do so), 15% medium-sized, and 5% large. In other words, the “average” represents a company which does not exist. This is only an illustration. Obviously, devious or incompetent agencies can abuse averages in other ways.

2.6 Accountability

All the best practice exemplars were accountable to stakeholders for the accuracy of their MS&A results, and none of the worst offenders seemed to have meaningful accountability. This also seems to hold true outside of government. This was tied for the best predictor of risk. Accountability is 82% correlated with satisfactory results and 42% correlated with good practice.

Accountability seemed to have at least three forms.

The first form of accountability: direct, repeated accountability to end-users of MS&A results. The EIA and UK Met Office publish their forecasts. Everyone knows when they are wrong. They are critiqued when they are wrong. Stakeholders complain when they are wrong. This led them to an emphasis on adopting best practices to improve their performance. Of course, their forecasts are still wrong at times. Best practices do not assure operational perfection. But this is a contrast to the worst offenders.

These “worst offender” organizations had all or most of the following attributes:

- Regulatory or statutory authority with few checks or balances

- Rare comparison of predictions with reality
- Restrictions on peer reviews
- Reliance on a certification, VV&A, or other endorsements
- Rotation of executive oversight
- Rejection of outsiders

These six “Rs” form a checklist to suggest risk factors for lack of direct, repeated accountability, which in turn seems to be the best predictor of MS&A pathologies.

It is worth noting that some of the best practitioners must deal with some of these six Rs. For example, some highly sensitive matters make peer reviews or including outside viewpoints very difficult. Best practitioners recognize these risks and find ways to lessen their impact.

The second form of accountability: clear knowledge of limitations. All the best in class or near best-in-class benchmarking participants had a clear understanding of their limits. This went beyond a defense of their budget or a request to grow resources. Rather, they could clearly explain why they faced limiting factors that either impacted their MS&A or created risk.

Part of accountability for exemplars was the duty to articulate the bounds of what they could not do. In contrast, the worst performers seemed quick to claim they could do anything and reject outside help.

The third form of accountability; lack inherent bias. This was an attribute of the exemplars. In contrast, inherent bias seemed a feature of most of the worst offenders.

Bias can appear in several forms:

- Power to pursue an agenda, supported by biased MS&A
- Power accrued to an organization or individual maintained by biased MS&A
- Pursuit of funding and resources by biased MS&A
- Confirmation of shared expectations, aspirations, and dogma
- This is not a complete list

Inherent bias seemed to be a feature of many Benefit/Cost Assessments (BCA) we reviewed. Many English-speaking nations have national laws requiring agencies to publish their estimate of the costs (and in some cases the benefits) of a new rule, law, or regulation. Some BCA estimates are breathtaking in the bias they seem to include.

The nature of human limitations is ripe with examples of subtle bias and blind spots. Addressing all bias may be too much to hope for. We only report here on what we think to be gross bias, which in many cases seems to be unchecked and unchallenged.

Bias is a good example of how pathologies can combine. Inherent bias, with authority lacking accountability, along with misuse of averages, is a very toxic combination. We saw this particular combination on more than one occasion.

3.0 ABOUT ARTIFICIAL INTELLIGENCE (AI)

Because AI has attained almost magical status, we wanted to understand the risks of automatically derived analytics. To do this, we conducted a separate exploration of risks in Artificial Intelligence. A full treatment of that work is not in the scope of this paper. Generally, we saw two classes of AI practitioners with different associated risks.

The first class of AI practitioners was trying to be “less wrong.” There are many problems where being even slightly less wrong can be unbelievably valuable. For example, product recommendations to online shoppers (Smith and Linden 2017) may be wrong most of the time, but still, generate billions of dollars of revenue for Amazon and others. Making recommendations at random is wrong essentially all the time. So, being wrong most of the time is a vast improvement.

For this first class of AI practice, the primary risk is that a decision maker relies too heavily on the recommendation. A human who asks you, “do you want fries with that?” is not responsible for your dietary choices. In the same way, these low precision recommender systems, and other AI systems that just aim to be less wrong, are offering hints for decision makers, not high confidence forecasts or predictions. When this reality is ignored, risk follows.

The second class of AI practitioners was attempting to make forecasts and predictions. They sought to do more than hints and being “less wrong.” Among this class who participated in our benchmarking, all exhibited serious errors in their work. We have been conducting follow-on research to understand why this seemed widespread. That work continues with the support and involvement of others. We hope to publish findings in the future.

4.0 SUMMARY

For NATO, the risk list is a means to improve performance, lessen the frequency of risk emergence, and lower the consequences when risk manifests.

Because NATO and its Partners deal with a wide range of topics, requirements, and threats, it might be wise to begin with consequences. Life and death consequences mandate a conservative approach. With lesser consequences at stake, implementing acceptable risk with the goal of being “less wrong” can provide agility.

Understanding risk and accepting it for the sake of speed is acceptable in some cases, while in others, avoidance of risk is paramount.

For MS&A practitioners, the risk list offers a means for self-assessment. These are the questions decision makers should be asking. As a matter of professional ethics and standards, practitioners should do a risk assessment even if not directed to do so.

Leadership and executives can encourage best practices and avoid consuming risky MS&A results by taking two actions. First, the risk checklist is an effective way to gauge how the risk in MS&A products is distributed across your organization. It was written for non-technical personnel and does not require practitioner-level knowledge to use. Second, we suggest asking your MS&A practitioners to use the work on best practices published in 2018, to conduct a self-assessment, and to suggest near-term actions for improvement.

A version of the checklist with the four most predictive topics is shown in Figure 5.

Assessing the Risk of Modelling, Simulation & Analysis (MS&A) used in an Organization

1. Accountability:

- a. Is the provider of MS&A accountable for accuracy and is anyone monitoring performance, such as accuracy?
- b. Is the provider shielded from accountability by any of the following?
 - i. Regulatory or statutory authority with few checks or balances?
 - ii. Rare comparisons with predictions of reality?
 - iii. Restrictions on peer reviews?
 - iv. Reliance on a certification, VV&A, or other endorsements?
 - v. Rotation of executive oversight?
 - vi. Rejection of outsiders?
- c. Can the MS&A provider describe any credible limits of accuracy and appropriate use?
- d. Is there reason to suspect bias (even if unintentional) has crept into the provider’s culture or delivered MS&A? Is there cause for concern if any of the following might be a source of the bias:
 - i. Power to pursue an agenda, supported by biased MS&A?
 - ii. Power accrued to an organization or individual maintained by biased MS&A?
 - iii. Pursuit of funding and resources by biased MS&A?
 - iv. Confirmation of shared expectations, aspirations, and dogma?
 - v. Any other reason to expect bias without accountable oversight?

2. Using single numbers as spans for proxies of uncertainty:

- a. Does the provider offer point predictions without any context of a range of possible outcomes or the odds associated with the number?
- b. Are averages used as proxies for all possibilities?

3. Are MS&A systems and tools static?

- a. Is there reliance on historical validation or certification?
- b. Are out-of-date methods like “p-value” confidence still in use?
- c. Is there reluctance to change from “our standard software”?
- d. Is there a dependence on supposedly irreplaceable gurus?

4. Is there a dedicated MS&A Team? Can they explain why/when other providers should be used?

Figure 5: Executive Risk Checklist for the Age of Algorithms

REFERENCES

- [1] Armstrong, J.S., Professor, Wharton School – University of Pennsylvania, personal communication, circa March 21, 2017.
- [2] Hoekstra, Auke (2018) “Photovoltaic growth: reality versus projections of the International Energy Agency – with 2018 update” available at <https://steinbuch.wordpress.com/2017/06/12/photovoltaic-growth-reality-versus-projections-of-the-international-energy-agency/>.
- [3] Petroleum Resources Management System, sponsored by Society of Petroleum Engineers (SPE), World Petroleum Council (WPC), American Association of Petroleum Geologists (AAPG).
- [4] Roemerma, S. et al., (2018). “Best Practices in Modelling and Simulation; Multi-Community Benchmarking”, Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL, paper 18125.
- [5] Smith, B. and Linden, G., (2017). “Two Decades of Recommender Systems at Amazon.com”, IEEE Internet Computing, vol. 21, no. 03, pp. 12-18.
- [6] Society of Petroleum Evaluation Engineers (SPEE), Society of Exploration Geophysicists (SEG), Society of Petrophysicists and Well Log Analysts (SPWLA), and the European Association of Geoscientists & Engineers (EAGE).
- [7] Wasserstein, R.L. & Lazar, N.A. (2016) “The ASA's Statement on p-Values: Context, Process, and Purpose”, *The American Statistician*, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108.
- [8] Worsley, P. M. (2009) “50 Years Ago: Cargo Cults of Melanesia”, *Scientific American* <https://www.scientificamerican.com/article/1959-cargo-cults-melanesia/>.
- [9] Zweidler, D. Professor, Senior Fellow, The Mack Institute for Innovation Management, The Wharton School - University of Pennsylvania, personal communication, circa February 10, 2017.

