Supplemental material to the paper:

Analytical problem solving based on causal, correlational and deductive models

Jeroen de Mast, Stefan H. Steiner, Wim P. M. Nuijten and Daniel Kapitan

A. The process of analytical problem solving

The paper discusses three frameworks for analytical problem solving: causal modelling, correlational modelling and deductive modelling. Table 3 in the paper summarizes the discussion by placing the three types of modelling in the process of analytical problem solving. In this supplemental material, we explain the structure of Table 3 and relate it to the literature.

Section 5.1 of the paper discusses essential differences between the three types of modelling. Table 3 organizes these in the problem-solving process. The literatures in various applied fields abound in models of the problem-solving process, typically organized in step-wise procedures. Table 4 collates four such procedures:

- Smith (1988), one of the best researched models in the management literature.
- Six Sigma's DMAIC model (De Mast and Lokkerbol 2012), widely used and discussed in quality engineering and statistics-driven operations improvement.
- The CRISP-DM model in data science (Chapman et al. 2000).
- PDCA and Toyota's A3 problem-solving procedure (Marksberry, Bustle and Clevinger 2011), which is representative for many similar problem-solving procedures in practitioner books.

Table 4. Four models for the process of problem solving

Smith	DMAIC	CRISP-DM	PDCA / A3
1. Goal setting	1. Define	1. Business understanding	1. Clarify problem
2. Description of current situation	2. Measure	2. Data understanding	2. Break down problem
3. Diagnosis (causal explanations)	3. Analyze	3. Data preparation	3. Set targets
4. Generation of alternative solutions	4. Improve	4. Modelling	4. Root cause analysis
5. Prediction of the effectiveness of solutions	5. Control	5. Evaluation	5. Develop countermeasures
6. Evaluation of alternative solutions		6. Deployment	6. See countermeasures through
7. Implementation and gaining acceptance for chosen solution.			7. Monitor results and processes
			8. Standardize successful project

Even though most of such models order problem-solving tasks in a numbered sequence of steps, they acknowledge almost universally that actual problem-solving processes are more iterative, with much going to and fro between tasks in cycles in which understanding of the problem becomes more and more refined.

Table 3 in the paper breaks down the task of finding improvements based on $Y = f(X)$ relations into six generic tasks, which are the rows of the table. Below, we explain each of these tasks in more detail.

1. Find Y ("What are relevant outcome variables?")
Driven by domain expertise and the problem owner's needs, the problem solver identifies measurable dependent (Y) variables that may be relevant for the problem. Typical challenges include clarifying implicit, vague or incoherent problem definitions and finding a reliable method for measuring Y. Typical techniques include problem structuring (Mingers and Rosenhead 2004), flowdown diagrams (De Koning and De Mast 2007) and measurement system assessment (Wheeler and Lyday 1989).

2. Find X ("What are relevant independent variables?")
The problem solver identifies candidate X variables that may be useful for predicting Y or understanding its behavior. This task echoes *Causal Diagnosis* in Smith's (1988) model (see Table 3) and *Root cause analysis* in the PDCA / A3 model, but identified X's can be any kind of predictors, causal or not. Some variables will be under the control of the problem solver (X^C), while other are uncontrollable variables (X^U). A typical challenge is avoiding tunnel vision, where the problem solver only identifies X's of a similar kind, and misses entire classes of potential predictors. Table 1 in the paper lists various approaches for identifying potential X's:

- Exploratory data analysis (unsupervised learning, graphical techniques, dimension reduction)
- Experiential knowledge (domain experts, case libraries)
- Sequence of hierarchical elimination studies (eliminate-and-home-in, branch-and-prune)
- Feature engineering
- Deduction from axiomatic theory

3. Find F_Y *and* F_U ("What is the current state?")
The problem solver investigates the current state and establishes the distributions $F_Y(y) = P(Y \leq y)$ and $F_U(x) = P(X^U \leq x)$ of the Y and X^U variables (including their means, variances, and other relevant statistical properties). A good characterization of the current state helps the problem solver refine her understanding of the problem and achieve focus by considering questions such as: Does the problem manifest itself in the mean or in the variance of Y? Does the problem evolve over time? Can the problem be broken down into subproblems, and which subproblem is the most severe? To what degree can the X's be controlled? Which X's have fixed values, and which have substantial variance? Knowing the means and variances of the Y and X^U variables also helps assess whether the problem owner's perception of the problem and desired end state are realistic. Typical techniques include descriptive statistics and estimation, the process-capability analysis (Wu, Pearn and Kotz 2009) and baseline studies (Steiner and MacKay 2005).

4. Find $Y = f(X)$ ("How are X and Y related?")
The problem solver builds a model that predicts Y (or F_Y) from the X's. This can be a causal, correlational or deductive model. The paper discusses suitable ways of building a predictive model:

Causal model	***Correlational model***	***Deductive model***
• Randomized controlled experiments • Structural causal modelling based on observational data	Observational study (supervised learning), based on representative sample from target population	• Deductive derivation from axiomatic theory • Approximation • Simulation

Typical challenges include causal inference, the bias/variance trade-off, and the analytical tractability of deductive models.

5. Find $V(Y, X^C, F_U)$ ("What is the objective?")
The problem solver translates the problem owner's objectives into a question about relationships between dependent variables Y, and controllable and uncontrollable independent variables X^C and X^U. The problem owner's desired end state is captured in a value function $V(Y, X^C, F_U)$ that is to be maximized. The problem solver also establishes constraints on the X's, and whether they should be treated as a controllable or uncontrollable variable.

6. Find $\max V(Y, X^C, F_U)$ ("How will we solve the problem?")
The various types of predictive models support various solution patterns, as listed in Table 2, which involve that the problem solver maximizes value for the problem owner by intervening in some of the controllable X^C variables. Interventions can be one-time optimizations (as in response-surface optimization) or continual dynamic adjustments as in predictive control.

Causal model	***Correlational model***	***Deductive model***
• Response-surface optimization • Robust design • Tolerance design • Decision optimization • Predictive control	• Decision optimization • Predictive control	• Response-surface optimization • Robust design • Tolerance design • Decision optimization • Predictive control

References

CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C.R. and WIRTH, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. The CRISP-DM Consortium.

DE KONING, H. AND DE MAST, J. (2007). The CTQ flowdown as a conceptual model of project objectives. *Quality Management Journal* **14**(2) 19–28.

DE MAST, J. and LOKKERBOL, J. (2012). An analysis of the Six Sigma DMAIC method from the perspective of problem solving. *International Journal of Production Economics* **139**(2) 604–614.

MARKSBERRY, P., BUSTLE, J. and CLEVINGER, J. (2011). Problem solving for managers: A mathematical investigation of Toyota's 8-step process. *Journal of Manufacturing Technology Management* **22**(7) 837–852.

MINGERS, J. AND ROSENHEAD, J. (2004). Problem structuring methods in action. *European Journal of Operational Research* **152** 530–554.

SMITH, G.F. (1988). Towards a heuristic theory of problem structuring. *Management Science* **34** 1489–1506.

STEINER, S.H. and MACKAY, R.J. (2005). *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes*. Quality Press: Milwaukee.

WHEELER, D.J. and LYDAY, R.W. (1989). *Evaluating the Measurement Process*, 2nd edition. SPC Press: Knoxville.

WU, C.W., PEARN, W.L. and KOTZ, S. (2009). An overview of theory and practice on process capability indices for quality assurance. *International Journal of Production Economics* **117**(2), 338–359.