Communicating a Model-Based Energy Performance Indicator

By Antony Hilliard, Greg A. Jamieson, & David Jorjani

To decarbonize the global energy supply will require technological, economic, and social adaptation. Human factors/ergonomics practitioners might anticipate barriers to these changes by studying contemporary energy management work, including the task of monitoring and targeting (M&T) energy use (Gotel, 1989).

Energy M&T has been adopted in industrial and commercial sectors, which together account for over 64% of global primary energy consumption (U.S. Energy Information Administration, 2013). However, hiring a dedicated energy manager is usually affordable only for consumers with annual utility (electric and gas) costs greater than about $2 million (Trombley, 2014).

This article describes why we found M&T work interesting, how we found gaps in existing information support, features of the design prototype that resulted, and where the project outputs are being applied.

INVESTIGATING M&T PRACTICE

M&T generates value by informing energy-efficient human behaviors, including investment, operation, and maintenance. Retrofit capital investments (e.g., more efficient heating) are easier to quantify, but if not operated and maintained mindfully, efficient systems can degrade (Mills & Mathew, 2009). M&T generates performance indicators that people can use to control energy use and discover ways to run their business more cost-effectively. However, despite the demonstrated benefits of M&T (Carbon Trust, 2008), it is not practiced universally, even among large energy consumers.

We collaborated with energy analytics company Energent Inc. (http://www.energent.com) to enhance its M&T software service. Energent wanted to discover unmet M&T work support needs shared by its clients and in-house analysts. We investigated M&T work and identified a poorly supported work need: communicating the energy models that underlie energy performance indicators. We next defined information requirements and designed a novel display wireframe.

Our main objective was to develop a useful concept that met our collaborator’s commercial needs, using cognitive ergonomics methods, such as

- participant observation,
- field observation,
- interviews,
- task analysis,
- wireframing, and
- tabletop evaluation.

We first investigated energy management by reviewing the literature and interviewing 18 energy management professionals (Hilliard, Jamieson, & White, 2009). They identified energy M&T as common but difficult. We found instructional M&T literature (e.g., Gotel, 1989; Hooke, Landry, & Hart, 2004) but no descriptions of M&T behavior “in the wild” (Hutchins, 1996).

Participant observation study. To familiarize ourselves with M&T work, the first author conducted participant observation. He worked as an M&T apprentice for 4 months at a large industrial manufacturer. The most prominent daily work challenges he observed were

- cultivating a meaningful data record,
- developing useful energy models,
IMPERFECT ENERGY MODELS

Models for energy monitoring and targeting (M&T) are developed to indicate “expected historical energy performance,” but defining expected and performance requires careful judgment. Regression statistics can be skewed by imperfectly linear, stable, and causal relationships between variables and energy use. They are also affected by the balance of system operation modes in the training data.

It is not that M&T models must be perfect – statistically flawed models can still be useful as long as workers interpret them correctly. An irony of M&T is that continuously improving energy performance conflicts with cultivating stable energy performance data. The more successful the work, the more difficult the performance indicators become to interpret. Even if analysts can collect sufficient measurements of steady performance, they may choose to omit variables or time ranges to tailor a model to a purpose.

For example, a building heating model that corrects for thermostat settings would isolate the effects of equipment efficiency. On the other hand, a model could encourage action (such as setting back temperature at night) by excluding thermostat settings, making the performance indicator reward behavior. Because the meaning of energy models can vary so greatly, an accessible model summary report is essential for effective M&T work.

- discovering energy savings opportunities, and
- communicating opportunities to colleagues.

These challenges informed additional M&T practice review and a task analysis of energy M&T, described in the following sections and sidebar.

Field observation and interviews. We next conducted a field study to corroborate our understanding of M&T work needs and to identify design opportunities (Hilliard & Jamieson, 2014). Three Energent client users and both in-house Energent energy analysts participated. Client users were a junior production engineer (Anna from Table 1), a facilities manager (Boris), and a maintenance supervisor (Bill). Energy analysts Carl and Chris supported their work.

We observed participants at their workplace and asked them to “think aloud” as they interpreted recent business energy consumption. Eight sessions averaged 20 min. We reviewed audio and video recordings of the participants’ work surfaces to categorize task behavior and corroborate our field notes. Opening and closing interviews addressed participants’ experience, findings of energy performance changes, and understanding of M&T methods.

DATA ANALYSIS

Analyses of the M&T practice literature, participant observation, and field study helped us identify M&T practice challenges, user goals, and concerns about existing M&T tools.

Practice challenges. A theme from our practitioner interviews was that energy management work is difficult because it requires “a technical person who can bridge multiple worlds in a generalist sort of way from a position of credibility” (Hilliard, Jamieson, & White, 2009, p. 3). Such a multitalented employee may not stay long at a salary that can be justified by typical 5% to 15% energy M&T savings. Conversely, an unskilled practitioner can wipe out savings with the labor costs of false alarms or mistaken diagnoses.

One solution to a skilled worker shortage is outsourcing. However, as a participant observer effectively acting as an outsourced M&T worker, the first author found that (a) the pace of business change meant that refurbishing out-of-date energy records took much of the work term, (b) only colleagues with local knowledge could judge the profitability

<table>
<thead>
<tr>
<th>Site</th>
<th>Participant Pseudonyms</th>
<th>Approximate Yearly Utility Consumption</th>
<th>Work Observation</th>
<th>Interviews</th>
<th>Task Walk-Through</th>
<th>Implementation</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client Site A: Light industrial</td>
<td>Anna</td>
<td>5 GWh electricity 8 Mm³ natural gas</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Client Site B: Institutional</td>
<td>Boris  Bill</td>
<td>25 GWh electricity 4 Mm³ natural gas</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborating energy management information system supplier</td>
<td>Carl Chris Cindy Connie</td>
<td></td>
<td>✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td>Changed employment Hired to replace Carl</td>
<td></td>
</tr>
</tbody>
</table>
Communicating a Model-Based Energy Performance Indicator

M&T task goals. The role of client champions was to interpret energy performance and motivate colleagues. Energy analysts supported this work by maintaining the data and models underlying M&T performance indicators. However, all participants expressed three M&T task goals: (a) judge whether energy performance was improving or worsening, (b) diagnose whether changes were practically significant, and (c) save money (by acting effectively).

To achieve these goals, participants relied heavily on a key performance indicator, presented as time-series cumulative sum (CUSUM) charts (Gotel, 1989). CUSUM charts show the difference between metered utility (energy) consumption and a model-predicted “expected” consumption.

Concerns with existing M&T tools. We identified model communication as a barrier to achieving M&T goals collaboratively using CUSUM charts. Of four M&T software tools we used during participant observation (two commercial products, one government-funded analysis tool, and one instructional spreadsheet), none presented comparative indicators of model quality. Existing features we observed (Figure 1) consisted of the following:

- Descriptive (but short) model titles
- Freeform “comment box” annotations
- Summary pie charts of model driver contributions
- Standard tables of summary statistics

The field study corroborated that existing software features were insufficient to diagnose model-based performance indicators. We observed that energy analysts and clients expressed different understanding of energy models (Hilliard & Jamieson, 2014). Energy analysts used offline statistical software outputs to assess models they created. Client users without this first-hand knowledge had a less reliable understanding of models, and none noticed week-to-week differences when energy models had been updated. Both Anna and Boris explained modeling processes as using more sophisticated data and methods than was the case. Analysts tried to cope with this mismatch by explaining energy models over telephone meetings but were concerned that this approach would not scale up.

It is hard to persuade people to take responsibility for performance indicators that they cannot diagnose, and workers who do not trust statistical technology will not adopt it (Ghazizadeh, Lee, & Boyle, 2011). Thus, we proposed to develop a report...
that conveyed model-relevant information (Figure 2) while improving on the status quo (Figure 1). Energent accepted.

**TASK ANALYSIS OF M&T**

To summarize field observations and systematically consider the information required for energy M&T, we developed a task analysis. We analyzed M&T in terms of high-level knowledge products and processes, so as to generalize across applications. It is summarized in Figure 2 as an information flow map (Rasmussen, Pejtersen, & Goodstein, 1994) of the standard M&T energy performance indicator strategy (Gotel, 1989).

The task analysis describes how the energy performance–indicating CUSUM chart (Figure 2, at top) is generated. Generating the chart requires energy models trained on historic metered utility energy consumption (e.g., electricity) and measures of conditions (such as production or weather) that drive energy use (Figure 2, bottom). When business conditions change, energy models must be maintained to reflect selected historic performance (Figure 2, center).

**Information requirements.** Most of the M&T work we observed involved interpreting CUSUM charts against past conditions or local context to distinguish influences on energy performance (see Mumaw, Roth, Vicente, & Burns, 2000). This interpretation requires skilled judgment and knowledge of influences on CUSUM charts, such as meter calibration (Figure 2, right), record-keeping policy (Figure 2, left), or unmeasured conditions. Furthermore, models can represent different aspects of historic energy performance (see sidebar). Resolving this ambiguity requires information about less-observable (darker-shaded) portions of Figure 2 to answer questions such as those in Table 2.

We determined that analyst and client field study participants could access different information to answer the questions in Table 2. Energy analysts could reason about model statistical quality but could only guess at the business conditions responsible for metered energy consumption. Conversely, client champions could explain week-to-week site conditions but misunderstood models and felt too busy to investigate.

**Figure 2.** Information flow in model-based energy monitoring and targeting. Measurements of the business system (bottom) are processed into an integrated, abstract performance indicator (top). Energy models are maintained and updated from time to time (center). Shading indicates how well documented and observable the information was. The report design focused on communicating elements in the darkest-shaded (least-observable) regions.
The information requirements in Table 2 clarified our design intent: Support evidence-based judgment about energy performance by communicating processes, products, and purposes that influence energy models. A shared artifact that presents models in accessible terms should reduce barriers and support effective, collaborative M&T by a wider range of practitioners.

**Design requirements.** The report design followed two principles: first, to serve as a unified collaborative reference for workers with backgrounds in management, finance, business analytics, and operations; second, to work in both online and printed formats. Although the design target was an interactive Web application, participant Anna argued that paper printouts better suited her needs for persuading production department colleagues. An interactive Web application can navigate large energy data sets, encourage decentralized use, and distribute up-to-date information. However, a design that also functions on paper becomes archival, reliable, easily shared, flexibly annotated, and more approachable for older users like Bill, who, like Anna, preferred printouts.

**WIREFRAME DEVELOPMENT**

We developed the report through three wireframe iterations. First, we generated a set of report elements to address information requirements from the task analysis (Figure 2 and the shaded questions in Table 2). We then arranged the elements to trade off

- a “pyramid” structure (summary first, information in priority, details last),
- a process flow structure (sequential or related model-training processes adjacent), and
- a fixed-format report structure to help navigate when comparing models (between revisions, purposes, or sites).

We next performed a heuristic review of the first prototype with four human factors graduate students. We used graduate students because they were outside our design team, complete novices to M&T, and familiar with using heuristics to discover usability problems (Nielsen, 2009). This review identified 22 suggestions – including unclear language, missing labels, and legibility – which we incorporated into a second prototype wireframe.

The second report prototype was evaluated by all Energent’s customer support energy analysts (Table 1). We conducted individual task walk-throughs with two report examples contrasting high and low model quality. The walk-through included semistructured questions, such as “Does...
the report suggest any weaknesses of this model?” or “Was the language in the report clear and understandable?” We noted which report elements participants referred to and how accurately they described properties of the model. Finally, the analysts and a customer-facing manager convened in a tabletop discussion of the report design.

We incorporated analysts’ suggestions and our observations into the third prototype wireframe, shown in Figure 3. For example, because one analyst did not notice the summary of data excluded from the model (see sidebar), we moved the content closer to the top left of the page and added a dedicated chart caption. Because all found the standards compliance term working range unclear, we substituted driver variable trained range.

**Wireframe design summary.** A conclusion from the tabletop discussion was to distribute the dense two-page format over an additional cover page to make it easier to introduce to novice clients. The cover serves to welcome users from a wide range of backgrounds and as an executive summary. The model is introduced in universal terms: the client business, site, application area, and historic range. Next, it is summarized by intended purpose, magnitude of utility consumption, associated costs, and share of internal or external influences.

The second page of Figure 3 describes how training data selection and historic energy consumption patterns (see Figure 2) reflect in the model. It explains model quality statistics in plain language and contrasts them with M&T standards compliance and model makers’ intentions. Finally, integrated time-series charts juxtapose familiar (e.g., business utility consumption) and model-specific diagnostic data. The final page describes the model by detailing the driver variables selected for the training set.

**EXAMPLE FROM ENERGY MODEL SUMMARY REPORT**

We present the most novel and complex feature of the M&T energy report in detail: the model training assessment chart (Figure 4). It describes the selected training set and model calibration process (Table 2) in an analogical graphic form with emergent features (Rasmussen et al., 1994; Wickens, 2013). Workers require information about model calibration because energy models can vary greatly depending on their assumptions. A model trained on hand-picked data, excluding unusual operating conditions such as a maintenance shutdown, may be more accurate during normal operation at the cost of responding misleadingly to nonroutine situations. Knowing what a model can represent is crucial to making good judgments based on its predictions (see sidebar).

The model-training chart integrates historic consumption, eliminated data, and model fit (shaded in Table 2). Upper charts in Figures 3 and 4 depict consumption during the training period. Distinctive business-specific patterns, such as seasonal heating and cooling cycles, can cue recall of training period conditions. The consumption chart also depicts abnormal patterns, such as flat lines from failed meters or dips caused by site shutdowns. Gray “excluded data” bands mark conditions that are not represented in the model.
During the tabletop discussion, analysts suggested overlapping model fit to actual energy consumption to demonstrate model-training accuracy. We implemented this suggestion (in Figure 4’s upper chart) as a de-annotated chart. We deemphasized this representation because the model-training error is shown only indirectly by the vertical gap between lines. This gap is hard to perceive and can be confounded with perpendicular distance due to the Poggendorff illusion (Wickens, 2013).

**Assessing models with CUSUM charts.** Model-training error can be displayed more consistently as a CUSUM chart, which serves as a “fingerprint” of model calibration (Zeileis, 2003). Four of our five field study participants used CUSUM charts in their day-to-day monitoring, and the recognition-based rules that workers learn to monitor CUSUM charts apply equally to assess model training.

We present three judgments about model quality that can be recognized using this chart, illustrated by comparing high-quality (Figure 3) and low-quality (Figure 4) heating models trained on real hospital data.

First, if hospital energy performance changes during the training period, the model will systematically over- or underestimate before and after the change, which creates phantom savings or losses. A CUSUM chart shows this problem as a distinctive $V$-shaped pattern above or below the horizontal axis. To emphasize and scale this stereotypical shape, we added symmetric guidelines of $\pm 2\%$ to frame the largest deviations from consistent fit. Both models show some sign of having been trained on unstable energy performance. Figure 3 suggests a small baseload improvement in mid-June; Figure 4, a moderate change of some kind in mid-March.

Second, when analysts exclude data from the training period, the resulting model will underrepresent conditions during those times. CUSUM chart shape can indicate whether data exclusions were reasonable. Chart roughness shows whether a modeler excluded data because of measurement time lag (jagged), error noise (spikes), or behavior unexplained by driver variables (a steep slope). The net effect of excluding data shows in the cumulative prediction error, which we labeled *training variance*. For example, if the October energy performance excluded from the model in Figure 4 was actually a routine annual event, the model would be expected to mislead by 1.6%.

Third, when “unusual” training period situations recur, the training CUSUM chart can serve as a template to interpret new M&T data. CUSUM charts of model training can help users confirm if the similar past events they recall produced recognizably similar performance changes.
Client adoption as a stand-alone tool. During the task walkthrough, all three Energent analysts agreed that they would find the report useful in their customer support and education work. They commented that the report contained all the information they would need to explain energy models to clients in a more clear and concise format than existing statistical software outputs. We did not conduct a quantitative evaluation, as Energent perceived an improvement over the state of the art and preferred to focus on implementing and deploying the product. This is not our first experience with a client directing limited resources to implementation over evaluation, and we expect that other practitioners have had similar experiences.

Energent has since implemented the model summary report as a stand-alone tool. Analysts are generating reports for clients as part of their support activities and revising the design based on client feedback. The product is slated to be included in the Web-based M&T software system in a future update. Future work should evaluate if model diagnostic information affects how clients interpret energy performance indicators.

FUTURE WORK IN PERFORMANCE MONITORING

Reporting tools for energy M&T work may benefit other domains in which performance is monitored using model-driven summary indicators. CUSUM statistics are a standard method in economics and quality control. They can be applied to monitor change in any time-series process that can be effectively modeled. For example, hospital-acquired infections are influenced by patient load, season, building age, and other less tractable factors. CUSUM performance indicator methods could be used to monitor the effects of human factors interventions, such as hand-washing or surface-sanitizing behavioral campaigns.

Energy M&T is part of a solution to climate change, given that "improvements in efficiency and conservation probably offer the greatest [carbon-reduction] potential" (Pacala, 2004, p. 969). Work-supporting tools are a crucial component of concerted, systematic efforts to speed adoption and effective use of modern energy technology and work practices.

However, existing M&T standards do not address tool ease of use (Efficiency Valuation Organization, 2012; ISO Technical Committee 242, 2011). In the residential domain, a draft Energy Star home climate control standard specifies ease-of-use requirements but has not yet been adopted (U.S. Environmental Protection Agency, 2014). Helping people understand and more efficiently control energy-intensive systems is one of many contributions that could be made by the human factors/ergonomics community.

REFERENCES


Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. Human Factors, 42, 36–55.


Antony Hilliard is a PhD candidate at the Cognitive Engineering Laboratory in the University of Toronto’s Department of Mechanical and Industrial Engineering. His research interests include cognitive strategies analysis as well as design of control interfaces, information systems, and physical devices.

Greg A. Jamieson is an associate professor and Clarice Chalmers Chair in Engineering Design in the Department of Mechanical and Industrial Engineering at the University of Toronto. He directs the Cognitive Engineering Lab, which conducts applied research in the natural resource and energy industries. He
specializes in human–automation interaction and the design of graphical interfaces for monitoring and control of safety-critical processes.

David Jorjani is a recent graduate of the Master of Information program at the University of Toronto, specializing in information systems and software engineering. He is interested in designing systems that facilitate smooth interactions among businesses, technology, and people. He is a technology consultant in the energy and utility industry.

Financial support for various aspects of this research was provided by the Federal Economic Development Agency for Southern Ontario, the Natural Science and Engineering Research Council of Canada, AITIA Analytics Inc., the Association of Major Power Consumers of Ontario, and the Ontario Centre of Excellence for Energy. The authors are particularly indebted to Energent Inc. for providing financial support, access to expert energy analysts and monitoring and targeting software developers, and opportunities to interact with its valued clients.

Copyright 2014 by Human Factors and Ergonomics Society. All rights reserved. DOI: 10.1177/1064804614550861