Recursive Estimates as an Extension to CUSUM-based Energy Monitoring & Targeting

Antony Hilliard, Greg A. Jamieson, University of Toronto, Canada

ABSTRACT

Statistical methods for energy consumption modeling have been static for the past three decades. Energy performance is typically modeled with linear regression and presented for interpretation using cumulative sum of residuals (CUSUM) charts. These mature techniques are general purpose, statistically robust, and simple. They suit energy measurement and verification (M&V) for quantifying a few large performance changes. However, in monitoring and targeting (M&T) applications, CUSUM charts are more challenging to interpret. Energy managers must disambiguate multiple known changes, detect unknown changes and, most importantly, diagnose and act on changes to energy performance.

We propose recursive estimates charts as a supplement to CUSUM charts to aid diagnosis in M&T. Recursive estimates charts track changes to linear regression model parameters. Performance changes correlated with model 'drivers' (e.g. more gas consumed during cold weather) are reflected as shifts in time-series charts. This provides informative diagnosis guidance to energy managers and can help them be more credible when engaging colleagues in correcting energy waste. We introduce an exponentially-weighted recursive estimate method modified to improve perceptual qualities, and demonstrate its application to an M&T case study.

Introduction

Energy monitoring & targeting (M&T) statistical methods have remained essentially unchanged since their development over 30 years ago (Technological Economics Research Unit, 1979). Following the oil crisis, British researchers adapted change detection statistical methods from econometrics and quality control (Brown, Durbin, & Evans, 1975; Page, 1961) to support energy management. These methods were adopted internationally, albeit with local variation. For example, early Japanese energy monitoring tended to be less centralized than British industry (Fawkes, 1986). Fawkes contrasted decentralized M&T, in which unsophisticated data was pushed to actively participating worker teams, against centralized M&T, in which data was accumulated for detailed analysis (often by a single expert).

These sociotechnical arrangements support different work strategies. Decentralized approaches encourage 'good housekeeping' and local work reorganization, while centralized approaches develop accurate and defensible evidence for capital investment and project accounting. This distinction will be familiar to any energy manager today although we are unaware of any accepted terminology reflecting it. We propose to retain the M&T label for monitoring strategies that emphasize controlling work activities (M&T) and to use measurement and verification (M&V) to refer to monitoring strategies directed at verifying the effects of planned investments. In statistical terms, both are *sequential change-point detection* problems in multivariate statistical process control. They differ in work purposes and the social environment that that guides action. Table 1 contrasts the two approaches.

Table 1. Contrasting Monitoring & Targeting, and Measurement and Verification Work

	M&T	M&V
Timescale	Ideally same as work activities	Accounting cycles
Focus	Detecting unplanned changes, diagnosing & correcting	Quantifying planned projects with known timeline
Accuracy	Should be informative. Depends on false alarm cost.	High enough to meet contract criteria or justify to investors
Purpose	Diagnose changes and improve system operation	Verify expected savings and investment returns

Of the two, M&V has been better described and standardized (Efficiency Valuation Organization, 2012), and its challenges seem to be mostly met by standard Cumulative Sum (CUSUM) methods. This paper will instead focus on challenges specific to M&T, which we will argue is not as well-supported.

The M&T Task and Missing User needs

A key purpose of M&T work is diagnosing energy performance changes. Unlike M&V work, quantifying changes isn't sufficient, and unlike classical quality control, M&T is applied in business systems that are loosely structured, open to outside influence, and adaptively reconfigured. Environments like this are subject to frequent change, and the challenge in M&T, like in other process monitoring tasks (Mumaw, Roth, Vicente, & Burns, 2000) is to efficiently filter sources of variability, identify patterns of interest, and diagnose changes. Changes need to be at least distinguished as being:

- Intentional or not? (Did energy performance change due to deliberate actions?)
- Actionable or not? (Can it be corrected, or only monitored?)
- Persistent or not? (Does it share the same 'root cause' as other changes?)
- Measured or Indirectly inferred? (Can its effects be easily modeled?)

Table 2. Four Useful Distinctions in Monitoring & Targeting Diagnosis

Example Energy Performance Change vs. Model	Intentional?	Actionable?	Persistent?	Measured?	Significance for Energy M&T Work
Seasonal Maintenance	X				Quantify if needed, Annotate, Ignore
Efficiency Retrofit	X		X		Suggests re-calibrating target model to correct for permanent change
Production Expansion	X		X	X	Suggests need to re-formulate target model with new driver data
Test Run	X	X			Determine if performance improved, report feedback to colleagues
Operations problem		X			Discuss with colleagues, diagnose, prevent recurrence
Maintenance problem		X	X		Record recurring effects to build case for repair
Control problem		X	X	X	Add measured data to preventative control strategy

The practical significance of a change, and the associated course of action, depends on such distinctions. Some examples are shown in Table 2. Statistical methods should support discrimination and diagnosis, or M&T will be more slow, costly, difficult, and ineffective.

The Social Environment of M&T

Statistical methods for decentralized applications also need to suit the social work environment. This was alluded to in the first comprehensive book on M&T:

> "There is an important parallel between money management and energy management. Information in a business usually flows in the form of … time series information ... However, the most fundamental indicators of business performance are not determined by analyzing financial information whilst it is in this form. The problem is then how to analyze it differently; so that it brings out the fundamentals which affect the performance of the business, but *retaining enough of the time series element to enable other people in the enterprise to relate it to their day to day activities*."(Harris, 1989, p. 7) (italics added)

This is a socio-statistical premise underlying M&T – statistics detect, colleagues diagnose. However, engaging colleagues carries social risks which we observed in a field study (Hilliard, Jamieson, & White, 2009). Motivating M&T work despite daily work pressures often requires "political capital" from a management champion, who gambles that energy savings will recover labor and opportunity costs. Fruitless searches make colleagues skeptical of the case for investigating energy waste, so proven success is needed to support decentralized participation.

These social needs suggest that statistical methods for supporting M&T need to be easy to explain to colleagues, match workers' understanding of the business, and support diagnosis to minimize problem search costs. Methods that induce incorrect diagnoses or false alarms can put an energy champion's personal credibility and program persistence at stake.

While CUSUM charts use easy-to-explain principles, and can be related to time-based understanding of business events, they don't offer much diagnosis support. We believe there is an opportunity for RE methods as a diagnosis aid to help effective, accepted decentralized M&T.

The CUSUM Method

We briefly explain the CUSUM method (Harris, 1989; Hooke, Landry, & Hart, 2003) so that we can contrast it with the RE method. A generic CUSUM task sequence is as follows:

1. Train a model to predict energy consumption **Y** based on measured independent variables $\mathbf{X_n}$ (aka. "energy drivers"). Linear regression is the most common modeling method.
2. For each time-series sample (e.g. daily) take the difference between actual energy consumption **Y** and driver-calcualted expected consumption at training-period performance **Y^** to calculate the residual error **ε**. Plot residuals as the control chart.
3. Integrate over the control chart residuals to calculate their Cumulative Sum, updated with each new data point. Plot this time series as the CUSUM chart.
4. Optionally, annotate the control chart with standard deviation-based limits, and the CUSUM chart with formal test statistic limits for a given decision criteria (Zeileis, 2003).
5. Interpret CUSUM chart inflection points (where the time series slope changes) as suspected energy performance change-points. Investigate these discrepancies.

This five step process understates the judgment needed to build models and interpret charts. We will focus on chart interpreting, as model building is beyond the scope of this paper.

Interpreting CUSUM Charts

The popularity of CUSUM charts is due in part to their match with human perceptual abilities. Perceiving small changes in a noisy control chart is difficult using scatter- or line-plots. A CUSUM chart helps by transforming the perceptual task from detecting change in mean position to change in angle (see Figure 1). In a monitoring CUSUM chart (Harris, 1989):

1. Straight lines correspond to periods of constant energy performance.
2. Sustained changes in CUSUM slope signify a change has occurred. The slope inflection point indicates when the change happened (or became observable).
3. Horizontal lines correspond to the energy performance of the model training period.
4. Jumps in the CUSUM plot correspond to one-time or brief energy performance changes.
5. Chart y-position shows "Cumulative savings/loss since the beginning of the period".

While these rules are sufficient for detecting changes in baseload, they are ambiguous, ineffective and potentially misleading when applied to slightly more complex changes.

CUSUM charts are easiest to interpret for baseload changes. Rules 1 & 2 above (straight lines = constant, “kinks” = change) are correct only for baseload changes. Intermittent energy performance changes and/or those associated with driver variables will produce fragmented slope changes, such as wavy “scalloped” patterns for seasonal heating/cooling variables (Harris, 1989, p. 23). For example, damage to a building envelope will increase utility consumption only in heating/cooling seasons. Thus, a single physical change to driver sensitivity might be misinterpreted using Rules 1 & 2 as several changes in baseload energy consumption.

Large baseload changes can obscure subsequent changes. If a large energy performance change occurs, such as major energy retrofit (savings) or a site expansion (overconsumption), the large model prediction residual error will be charted as a steep CUSUM slope. This is useful for quantifying changes as part of M&V, but for M&T purposes the effects are un-actionable and should be ignored. However, models can’t be immediately updated to correct for “new normal” energy performance, as a year’s worth of data is recommended for standards-compliance (Efficiency Valuation Organization, 2012).

Meanwhile, the CUSUM chart will tend to form a diagonal line, which obscures subsequent, smaller energy performance changes in two ways. First, it expands the vertical axis and compresses the chart scale. Second, humans are worse at perceiving changes between obliquely angled lines than between a horizontal line and an angled line (Appelle, 1972). In such cases, a control chart vertical shift may be a better indicator after all.

Overlapping changes produce ambiguous CUSUM plots. CUSUM plots aren’t very informative in distinguishing overlapping changes, even if changes don’t obscure each other. For example, if consumption is shifted from one period to another (e.g. weekday to weekend), or if a driver-related efficiency change is offset by a coincident change to baseload, a CUSUM chart will show only the net difference. This difference will vary with the interaction of the two variables and its start/stop times may mislead and suggest unrelated events.

Recursive estimates charts are a novel alternative, using familiar time-series graphs to resolve ambiguities in CUSUM charts.

Recursive Estimates as a CUSUM Supplement

The recursive estimates (RE) class of structural change detection methods were invented after the introduction of energy MT&R (Ploberger, Krämer, & Kontrus, 1989) and have been refined (Kuan & Chen, 1994), implemented in statistical software (Zeileis, 2003), and remain a common econometric analysis tool. Similar to CUSUM charts, RE charts detect system changes based on model residual error, producing a time series empirical fluctuation process. RE and CUSUM "are thus based on identical ingredients, which are however put to different uses. In fact, the [recursive estimates] Fluctuation test can be viewed as a backward CUSUM type procedure." (Ploberger et al., 1989, p. 312).

RE charts use model error to estimate *how changes are explained by model parameters*. In M&T applications, this can help diagnose energy performance in terms of processes represented by driver variables (e.g. heating efficiency). A task sequence for RE M&T is:

1. Train a linear regression baseline model, as for a CUSUM approach. Save two model training properties: the 'historic' model coefficients $\boldsymbol{\beta}_{\mathbf{n}}$, and the normalized covariance matrix of the driver variables $\mathbf{Q}^{1/2}$ (Ploberger et al., 1989).
2. For each time-series sample **i** following the training period, append the new data to a monitoring set and re-fit the linear model to all data up to time **i**. Calculate the difference between recursively fit monitoring model parameters and historic values ($\hat{\boldsymbol{\beta}}_{\mathbf{ni}} - \boldsymbol{\beta}_{\mathbf{n}}$).
3. Transform each resulting parameter change sample into an RE chart point through vector multiplying by the historic covariance matrix $\mathbf{Q}^{1/2}$. This corrects for normal inter-parameter coupling present in the training data. One way to think of $\mathbf{Q}^{1/2}$ multiplication is as a coordinate transform, from the "state space" defined by the driver variables into a new "change detection space" that is more orthogonal to (and thus less affected by) usual covariance between driver variables.
4. Normalize the scale of the RE charts by dividing by each parameter's training variance and a ratio of the monitoring time elapsed to the length of the training period. This produces unitless values compatible with test statistic limits (Kuan & Chen, 1994).
5. Interpret the resulting RE charts. Inflection points mark times where a performance change may have occurred. If using test statistics, note if/when the RE chart exceeds the test statistic limits, and conclude a change has happened in that model parameter.
6. For suspected parameter changes, consider what the model parameter represents in the real energy-consuming system and infer where to search for related causes.

Many of these steps are algorithmic and described more completely elsewhere (Zeileis, 2003). We will focus on judgment rules for interpreting RE charts.

Interpreting RE Charts

RE charts in the canonical form described above have been developed for data exploration and formal hypothesis testing (Zeileis, 2003). A RE chart transforms the diagnosis task of examining residual scatterplots into a CUSUM-like perceptual task (see Figure 2):

1. An RE chart fluctuating around zero indicates a model parameter with similar energy performance as the model training period.
2. A sustained change in position signifies a parameter change may have occurred.
3. Sharp jumps in the RE plot correspond to intermittent changes in energy performance.
4. If using formal test statistics, an RE plot that exceeds the test statistic envelope indicates when a change can be judged statistically significant.

RE charts complement CUSUM charts by partitioning the residual error into contributions of each variable in the energy model. This can support M&T by helping distinguish changes in energy driver sensitivity from baseload, disaggregate simultaneous changes, and initiate diagnostic searches. RE algorithms use the same linear regression functions familiar to CUSUM practitioners (aside from the novel $\mathbf{Q}^{1/2}$ scaling). However, using regression in this way introduces some statistical weaknesses and perceptual ambiguities that must be considered:

Statistical weaknesses of RE charts. RE charts depend on good models trained with driver variables that are independent, timely, accurate, and complete (as recommended by e.g. Efficiency Valuation Organization, 2012).

- RE methods can't outperform their drivers. If a system change is ambiguously shown in driver variables, the estimate fluctuations will be vague or indistinguishable.
- Model Autocorrelation, caused for example by time delay between consumption and driver variable measurement, distorts RE chart scale (Kuan & Chen, 1994)
- Unmodeled aspects of the energy-consuming system, such as measurement error or missing driver variables, affect RE charts more than CUSUMs (Young, 2011, p. 42). This is called ‘endogeneity’ and has three main effects:
- Regression Dilution: Measurement noise in driver variables will distort parameter estimates, usually underestimating parameters and overestimating baseload.
- Regression Contamination: RE charts aren’t perfectly independent, even when transformed into $Q^{1/2}$ ’coordinate space’. Measurement error, effects of un-modeled drivers, and even single parameter changes will contaminate parameter estimates for other variables according to their covariance.
- Test statistic inflation: Changes during training obscure later detection (Perron, 2006).

These RE flaws are either inherited from linear regression or encountered in applying statistical tests. CUSUM charts are more resilient to these flaws, and therefore will be more reliable in detecting changes. However, even if statistically imperfect, RE charts can still be useful in diagnosis if their shape clearly illustrates informative time-series phenomena.

Perceptual ambiguities in RE charts. RE charts in their standard implementation have properties that complicate interpretation.

- The RE algorithm recursively grows its monitoring set with each new data point, so charts grow sluggish and don’t distinguish intermittent and persistent changes.
- Straight-line RE charts can result from a driver variable holding constant for long periods (such as seasonal temperature drivers). This means that if a physical system change

occurs while a driver is 'stagnant', the effect won't be charted until the associated driver varies again. This ambiguity encourages misinterpreting change dates.

- Constant energy performance does not necessarily produce straight-line RE charts, due to chart scaling. Interpreting using CUSUM rules-of-thumb could lead to false positives.
- Scaling and coordinate transforms in RE abandon the engineering units of regression models, so unlike CUSUM charts, RE axis values don't have direct physical significance.

These challenges to interpreting RE charts exist in part because the algorithm has been developed to match the test statistic, and the test statistic is defined in terms of statistical certainty over the complete time history. But, as discussed below, test statistics may not be especially useful in an M&T task environment. Modifying RE to remove the test statistic may allow the charts to be more easily interpreted in an M&T context.

Implementing a Modified RE Algorithm

In Energy M&T, ease of interpretation by informed colleagues is more important than statistical certainty. Removing the test statistic from the RE charts allows some perceptual ambiguities to be designed out. To a statistician, discarding formal tests may seem foolish. Test statistics are necessary when risky decisions must be made solely on well-understood data. However, tests are of limited use in an M&T work environment:

- The pace of change in most businesses will tend to produce multiple overlapping changes, some of which are non-actionable. If test statistics are triggered by known changes, findings will be obvious and unhelpful.
- Machine-readable driver data is limited, so incorporating or correcting for known changes takes time and adds cost. Test statistics can't be tailored to every hypothesis.
- Statistical assumptions may not hold, making test statistics misleading (because of endogeneity such as measurement error, missing variables, etc.).
- Human-readable data is available. Colleagues can investigate to gain more knowledge on which to base a decision.
- Test statistic calculations and interpretations must be explained to colleagues, and may not help to inform or convince them.
- Calculations that can't be performed with commodity office tools (e.g. spreadsheets) may not be widely adopted or included in training.

Such objections may explain why existing statistical tests for CUSUM methods (Zeileis, 2003) aren't used in M&T practice.

Modifying the RE Algorithm

Removing test statistics won't fix statistical flaws of the RE method, but it allows three key improvements to be made to the task-relevance and perceptual features of the RE charts.

Exponentially weighted memory. Normal RE charts recursively add new data to an ever-growing monitoring set. This causes the RE chart to behave differently over time (as more data is recursively added), and averages out the characteristic time-series shape of intermittent changes.

To make RE charts more consistent and responsive, the monitoring time-series set can be weighted by an exponential decay (Young, 2011), so the linear regression will respond more quickly to 'fresh' samples and "forget" old data. The decay rate can be tuned to pick out phenomena of interest, analogous to adjusting a low-pass filter. This can be done either interactively in an on-line application, or for pencil-and-paper use, short and long "memory" RE charts can be plotted simultaneously.

Figure 2 shows both 2-month and 5-year time constant RE charts. They are formatted differently to distinguish them and indicate their behavior. The 5-year RE chart behaves similarly to the CUSUM chart, tracking persistent shifts in performance. The 2-month decay chart behaves similarly to the control chart, depicting abrupt changes as well as uncontrollable variation.

Meta-information. Intermittent drivers are commonly used in M&T models. For example, seasonal heating/cooling factors are inactive half the year, and products may be manufactured rarely. While a driver is constant (zero, for example), nothing can be learned about its influence on energy performance (you can't tell the furnace is broken until you turn it on). Therefore during such periods, an RE chart should be de-emphasized since it's not informative.

A simple metric for the informativeness of an RE chart point is relative driver variation, which can be calculated as the ratio of variation in recent monitoring data to that during model training. This metric is used in Figure 2 to vary the RE chart's transparency so that it fades away as the parameter estimate becomes 'stale'. The same transparency effect highlights missing data.

Relative scaling. A final RE chart design choice is axis units. Since estimated parameters aren't always precise due to statistical flaws of the RE algorithm, axis units should encourage qualitative exploratory diagnosis rather than quantitative M&V use. One option is to normalize each parameter deviation by the historic trained model parameter $\boldsymbol{\beta}_n$ *before* $Q^{1/2}$ *transformation.* This converts to a "percentage change" unit, which while still without direct physical significance at least indicates the order of magnitude of a possible change.

Interpreting Modified RE Charts

Specifying interpretation rules for a new statistical application is a bit imprudent, as different work environments will produce patterns that local experts will learn to recognize with practice. However, for instructing novices, we find these rules a good starting point:

1. Identify change times with the CUSUM chart. Ignore known/intentional changes.
2. For each remaining unexplained change, check for *simultaneous* RE chart changes.
    1. Use long memory RE charts to see which parameter chart "most resembles" the CUSUM at the change time.
    2. Use short memory RE charts to check if parameters also have a spike or step-change at the change time.
3. If one parameter's RE chart seems consistent with the CUSUM change, use short memory RE charts to determine whether the change persisted or reverted.
    1. If the change persists before or after, note other CUSUM changes that may be due to the same 'root cause' (such as poor performance in rarely-used equipment). Go investigate, diagnose, and take action.

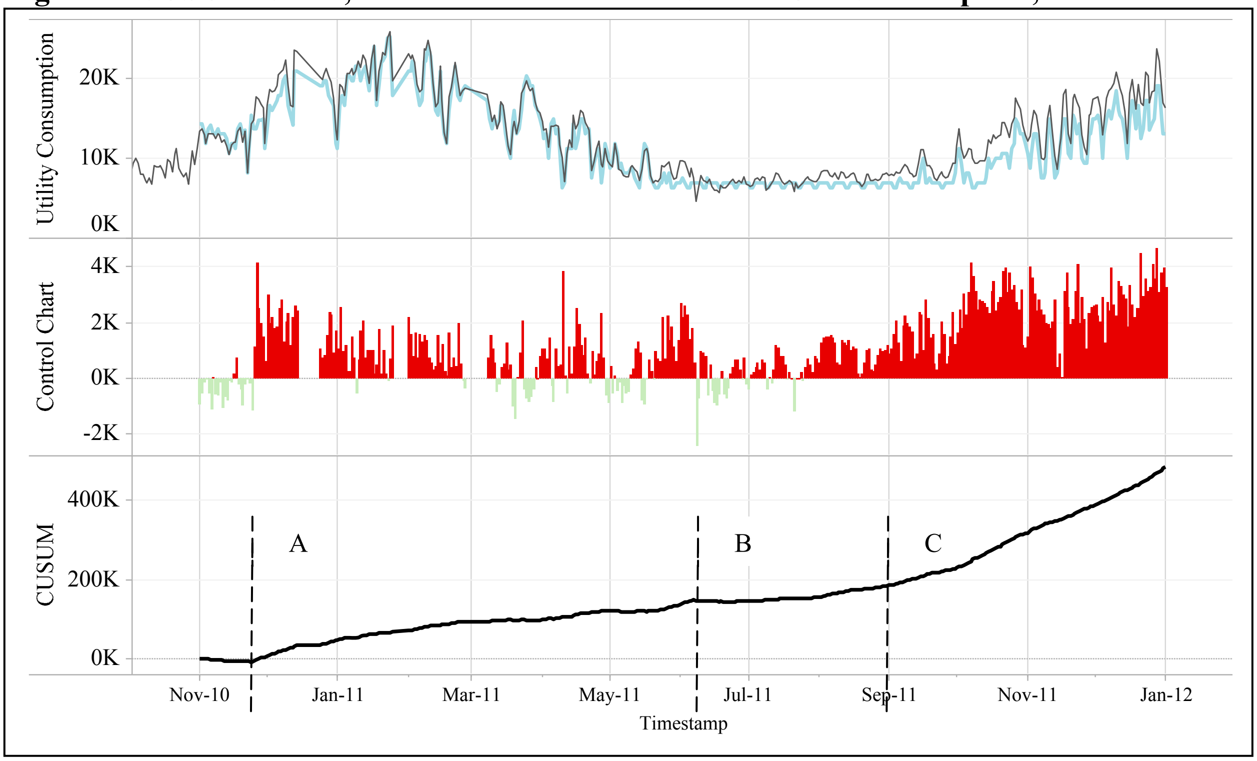
2. If the change reverted, note the date it reverted, ensure it is consistent with the CUSUM, and investigate historic events consistent with the start/stop dates.

A case study will demonstrate applying these rules to real data.

Case Study: Healthcare Institution

We conclude this paper by demonstrating our modified RE method on actual data from a healthcare institution. The "base truth" of the changes discussed in this case study have been confirmed, but as with any real utility data, other un-investigated changes are also present. This Natural Gas consumption model was trained on 350 data points over 1 year, from November 2009-2010, correcting for Weekday/Weekend variation and Heating demand. It is fairly high quality, with a R^2 of 0.96, and a CV_{RMSE} of 8.7%.

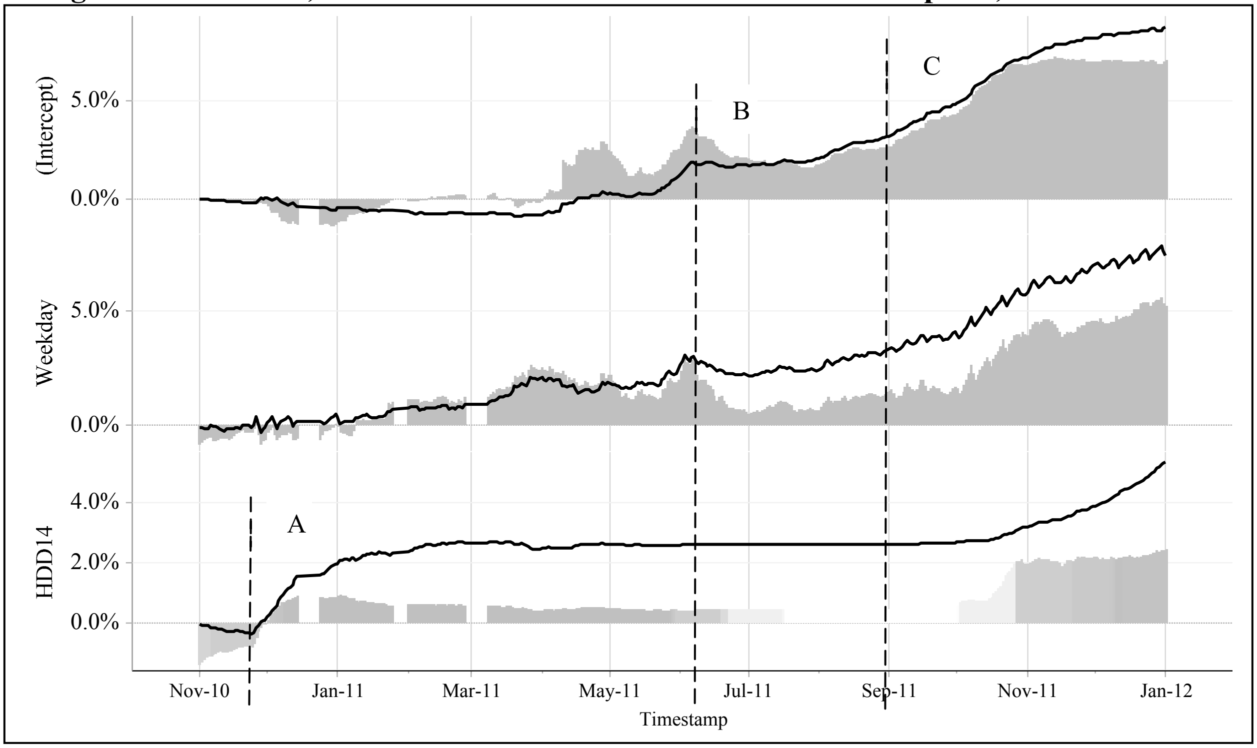
Figure 1. CUSUM Chart, Healthcare Institution Natural Gas Consumption, Fall 2010-2011



From top to bottom: Consumption (black) and Modeled consumption (blue) charts. Control chart showing over/underconsumption. Finally, CUSUM chart showing times of three suspected changes (A,B,C)

Interpreting the site CUSUM chart in Figure 1 might suggest three changes – One in late November 2010 (A), which moderates by June 2011 (B), and a second, larger change around September 2011 (C). The reader can try diagnosis using the CUSUM chart before turning the page.

Figure 2. RE Chart, Healthcare Institution Natural Gas Consumption, Fall 2010-2011



RE charts for the same model and period as Figure 1. Each chart indicates change in a model parameter: baseload (Intercept), Weekday, and Heating Degree Day. Each chart shows two exponential memory decay factors: Grey bars indicate a fast-responding, 2-month time constant. Black lines indicate a slow-responding 5-year time constant.

The three changes identified from the CUSUM chart are marked on Figure 2. 'Change A' corresponds to an abrupt change in the Heating Degree Day (HDD) parameter on November 24^{th}. The short-memory RE chart shows that the heating sensitivity moderated after January 2011, but did not completely revert. The abrupt increase is consistent with known events at the hospital: a contagious disease outbreak required the heating system be changed to 100% fresh air, temporarily increasing the building's sensitivity to cold weather. It is possible that the heating controls were not reverted completely after the outbreak was contained.

At the time of Change B, charts of Intercept and Weekday parameters had been increasing, indicating gas consumption unrelated to outdoor temperature. Both decrease somewhat as the (stable) heating parameter "fades out" during the summer. This change likely indicates inefficient boiler operation at the end of the heating season. Note that the brief "blips" are more noticeable on the RE charts than the CUSUM chart.

Change C can be distinguished as three separate events – an increase in the Baseload intercept over September-October, followed by increased Weekday consumption, and finally an increase in the Heating parameter in late October. This is consistent with the commissioning of a new building wing, whose interior was being finished during the fall. The unoccupied space may have required heating at a higher temperature break-point in October, but over the winter the increased heating load seems quite stable.

This example demonstrates how RE charts can help diagnose CUSUM changes. These changes are only those that could be verified after-the-fact, but discovering unplanned changes on-line follows similar rules, with subsequent search in real-time data and by colleagues.

Discussion and Conclusion

Recursive estimates charts are a modest extension to CUSUM charts, intended as a diagnosis support aid. They complement CUSUM charts, being less statistically robust, but more informative, while remaining reasonably easy to interpret. They also complement energy sub-metering by helping direct diagnostic searches by the functional units of driver variables.

We hope this proposal spurs development in M&T methods. Reviewing a classic list of M&T challenges (Fawkes, 1988), it is curious how many still present difficulties today, e.g.:

- Distinguishing controllable from uncontrollable energy consumption variability
- Proving negatives – "Convincing people that they have actually saved energy when they have actually used more, and spent more, can be difficult" (Fawkes, 1988, p. 312)
- Framing energy performance in terms that managers will accept responsibility for
- Over-spending on technology that doesn't meet information needs

We hope RE methods will be successful in distinguishing energy consumption variability, communicating energy performance to colleagues, reducing the need for sub-metering, and allowing inexpensive data to be interpreted to its full potential. Future work will be required to achieve these goals.

Psychological studies of M&T statistical methods and graphics are missing from the literature. An analysis of the cognitive strategies people use to understand 'normal' and 'surprising' patterns in energy data may help better understand skilled performance, communicate M&T findings and teach the next generation of practitioners (Bobker, 2004).

Modeling strategies are another opportunity for development. A universal problem in empirical change detection is distinguishing whether deviations are explained by system performance, or data/model quality. RE charts will only be useful if colleagues can understand what model parameters represent. For example, subtle differences in how driver variables are measured (such as the difference between pre- or post-scrap production quantities) can change how a diagnostic search should proceed.

Driver variables may be chosen to imbue the model with a designer's purpose. In M&T models, some controllable variables (e.g. scrap production) may be deliberately omitted in order to "reward" behavior that reduces the variables. Controllable variables may also be omitted because they aren't easily represented in linear regression (such as Heating Degree Day breakpoint temperatures). Consideration is needed of how model design affects interpretation.

Finally, the M&T community must address an irony of modeling: as an organization becomes more active in improving energy performance, long training periods with stable energy performance can't be found, and model quality decreases. One possibility to compensate for this effect is communicating which phenomena a model "knows", and which it will have difficulty correcting for. Both RE and CUSUM charts can be applied to identify if and when changes occurred during a model training period (Zeileis, 2003). Communicating these properties is an opportunity for future research.

Acknowledgements

Data courtesy of Energent Incorporated, who supported this research along with the Natural Sciences and Engineering Research Council of Canada and Federal Development Agency for Southern Ontario.

References

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals. *Psychological bulletin*, *78*(4), 266–278.

Bobker, M. (2004). Knowledge Practice in a Sea of Information. In B. L. Capehart (Ed.), *Information technology for energy managers* (pp. 171–178). Lilburn, Ga.: Fairmont Press.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society*, (35), 149–192.

Efficiency Valuation Organization. (2012). *International Performance Measurement & Verification Protocol: Concepts and Options for Determining Energy and Water Savings* (7th ed., Vols. 1-III, Vol. I). Efficiency Valuation Organization. Retrieved from www.ipmvp.org

Fawkes, S. D. (1986). A comparison of British and Japanese industrial energy management. *R&D Management*, *16*(4), 309–316. doi:10.1111/j.1467-9310.1986.tb01177.x

Fawkes, S. D. (1988). Monitoring and targeting in buildings: Principles, programs and processors. *Applied Energy*, *29*(4), 307–317. doi:10.1016/0306-2619(88)90041-4

Harris, P. (1989). *Energy monitoring and target setting using CUSUM.* Cheriton Technology Publications.

Hilliard, A., Jamieson, G. A., & White, A. (2009). *Energy Management in Large Enterprises: A Field Study* (Technical Report No. CEL09-01). Toronto, Canada: Cognitive Engineering Laboratory. Retrieved from http://cel.mie.utoronto.ca

Hooke, J. H., Landry, B. J., & Hart, D. (2003). Energy management information systems - Achieving Improved Energy Efficiency: A handbook for managers, engineers and operational staff. Retrieved from http://oee.nrcan.gc.ca/publications/industrial/EMIS/index.cfm?attr=24

Kuan, C.-M., & Chen, M.-Y. (1994). Implementing the fluctuation and moving-estimates tests in dynamic econometric models. *Economics Letters*, *44*, 235.

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*(1), 36–55.

Page, E. S. (1961). Cumulative Sum Charts. *Technometrics*, *3*(1), 1–9.

Perron, P. (2006). Dealing with Structural Breaks. In *Palgrave handbook of econometrics. 1, Econometric theory.* (pp. 278–353). Basingstoke: Palgrave Macmillan.

Ploberger, W., Krämer, W., & Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, *40*(2), 307–318. doi:10.1016/0304-4076(89)90087-0

Technological Economics Research Unit. (1979). *A study of the feasibility of energy costing and energy accounting models for management* (No. 15). University of Stirling.

Young, P. C. (2011). *Recursive estimation and time-series analysis.* Berlin: Springer-Verlag.

Zeileis, A. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, *44*(1-2), 109–123. doi:10.1016/S0167-9473(03)00030-6