Gold Mining in Your Back Yard: Discovering, Exploiting and Measuring Unsuspected O&M Savings Opportunities with Diagnostic Benchmarking: Field Trial Results

Les Lambert, Lambert Engineering

ABSTRACT

$150 billion/yr. in potential US energy savings, enhanced competitiveness, and reduced CO2 emissions depend on swiftly adopting Diagnostic Benchmarking. Expected savings consist mostly of no-cost Operating and Maintenance ("O&M") improvements in existing buildings.

Early energy audits required that O&M be improved before considering retrofits. But audits didn't effectively identify O&M opportunities. Appropriate measurement and verification, to ensure O&M problems got fixed, and stayed fixed, wasn't available yet. Auditors and contractors, perceiving O&M problems as difficult and unprofitable, relative to retrofits, gave O&M little attention. Management saw O&M as overhead to be minimized.

"Universal Benchmarking" recently showed early O&M problems haven't disappeared. Apparently similar buildings differ greatly in annual energy use intensity (EUI). Newer or larger buildings sometimes show relatively high EUIs. Universal benchmarking implies large O&M differences between buildings, but doesn't explicitly reveal O&M problems.

Concurrently, Diagnostic Benchmarking in 31 buildings achieved substantial savings. Despite uncompleted O&M and retrofits, no-cost O&M savings equaled 18.7% of initial energy use. Diagnostic Benchmarking of individual buildings proved essential to quickly find, capture, ***measure*** and retain O&M savings, and enabled important findings:

- For triage, higher fuel-specific seasonal EUI predicted greater savings; 100% of measured savings came from 60% of benchmarked buildings.
- Higher seasonal EUI or gas use, especially in newer buildings, and smaller size, indicated probable higher O&M savings percentages.
- Continuous benchmarking quickly detected non-obvious failures; correcting such failures generated significant savings.
- Benchmarking enabled right-sizing of replacement equipment.

Measured savings trends suggest new strategies and infrastructure for building efficiency.

Introduction

US buildings use over $500 billion/yr. in energy (NYC 2012). If 30 % of this energy were quickly saved at little or no up-front cost, without government financing, the impact would be roughly equivalent to a $150 billion/yr. tax cut, and an exemplary case of "reaching for high energy savings". Obvious questions are: Are such savings possible? If so, how, and why are we not doing so? How much of this $150 billion/yr. can be achieved through O&M improvements? What changes are needed to achieve these savings? This paper examines process observations and measured O&M savings from using Diagnostic Benchmarking (Lambert 2012b) in a field trial and recommends strategic and infrastructure changes to begin to achieve such savings.

In the author's experience, no practical benchmarks were available in the early 1980s to quickly indicate whether O&M showed lasting improvement, or if buildings performed as expected. Audits were partly aimed at finding and correcting O&M deficiencies, but O&M savings were difficult to find, verify, and retain. Much of the author's monitoring work was to demonstrate efficiency features in new buildings, but performance sometimes fell short of expectations due to O&M deficiencies. "Twin studies" found that physically identical buildings showed surprisingly large unrecognized performance differences due to different O&M (Lambert Engineering 1991, Lambert 1998). Yet most energy efficiency efforts continued to focus on design, new construction, and retrofits well-suited for replication. Few in the energy efficiency community seemed aware of the large adverse impact flawed O&M often has on efficiency.

In 1984, the author started developing benchmarks to solve these problems. The field trial reported here is the result of those efforts. As the field trial progressed, it became obvious that diagnostic benchmarking worked very well, and that its rapid introduction was crucial to reaching for high energy savings. This paper is the first published impact evaluation rigorously showing the benefits and large energy savings potential of Diagnostic Benchmarks.

Diagnostic Benchmarks use EUI in weekly average BTU/Hr-Ft^2, for each fuel. Alternative benchmarks use annual EUI, for all fuels combined. Weekly EUI, plotted or trended versus coincident outside temperature, reveals how building site energy use rates vary seasonally. By using weekly averages of site energy data, these curves avoid significant sampling errors inherent in benchmarking methods using monthly or source energy data. They are more statistically robust, providing the fastest, most accurate characterizations of whole-building energy use patterns possible. The curves and associated trends remain constant unless the building's physical plant or its operation changes. The effects of deliberate or unintended changes can be measured by the amount they shift the curves. Unlike alternative benchmarks, diagnostic benchmarks are intrinsic to the building and its operation, independent of site climate, source energy carbon intensity, or peer group membership; they are completely objective.

Diagnostic benchmarks are diagnostic in several ways. Comparisons among similar buildings indicate which will likely show the greatest savings from improved plant or operation. Patterns of energy use versus outside temperature can indicate unneeded heating or cooling energy use. Keeping fuels separate helps to understand interactive effects, and shows which fuels offer best savings potential. For individual buildings, an increase in seasonal EUI indicates an equipment fault or an adverse change in O&M that needs attention. Most importantly, a decrease in EUI provides an immediate indication of the merit of a change, and offers a means to measure the savings from either plant or O&M improvements. Its near-real-time information and unsurpassed accuracy enables finding, capturing, ***measuring*** (not previously possible), and retaining O&M savings. As audit precursors, diagnostic benchmarks transform audits from cookie-cutter exercises to focused investigations, done only when needed. See (Lambert 2012b) for more theory, details and building case studies using Diagnostic Benchmarking.

Why Are We Not Already Doing This?

The author's experience is that ineffective measurements have left stakeholders oblivious to wasteful O&M practices, and how to reliably manage them. Other diagnostic benchmark uses, such as commissioning, prioritizing work or retaining savings aren't yet recognized. O&M is widely viewed as an overhead cost to be minimized, instead of optimized. O&M staff often equate comfort and minimal downtime with success. Obvious problems that do not degrade comfort get scant attention, and invisible causes of waste are out of sight and out of mind.

The Environmental Protection Agency (EPA) Energy Star Portfolio Manager (ESPM) is becoming a de-facto standard for building benchmarking (EPA 2013). But ESPM wasn't designed for finding or measuring O&M opportunities. ESPM's current form has subjective elements, and is neither precise nor responsive enough for near-real-time measurement of O&M effects or changes to O&M.

Savings Measurement and Verification (M&V) has been imprecise and expensive, and is sometimes characterized as a diversion of funds from needed retrofits. Some view effective low-cost or no-cost M&V as a threat to their livelihood. Diagnostic Benchmarking must compete for priority with many product or service-oriented savings strategies, and their well-entrenched special interests. Major beneficiaries of Diagnostic Benchmarking- building owner/operators and policymakers - need to become aware of its uses and benefits. A field trial was in order, to demonstrate uses and benefits of Diagnostic Benchmarking, and to begin to explore process issues relevant to its introduction.

What Is a Reasonable Level of O&M Savings to Expect?

Diagnostic Benchmarking, early in the field trial, showed rough estimated savings of 32%. These savings were from one group of buildings; about half the savings were from O&M improvements. Two questions are, can we carefully refine the savings just from completed O&M, and are these buildings reasonably representative of the general building stock?

Determining final savings from field trials depended on unscheduled completion of remaining work - O&M improvements identified but not completed, and several retrofits. Rather than wait for that, we chose to carefully measure the savings from the O&M measures that ***were*** accomplished, to establish a minimum value for what was possible in this group of buildings.

Available data doesn't support discussing individual measurements of our retrofit savings. Most retrofits proposed in the field trial were intended to enable buildings to ***operate*** more efficiently, usually by means of variable frequency drives (VFDs). However, we have existing data that enables directly measuring the O&M savings accomplished using Diagnostic Benchmarking in the field trial.

The measured O&M savings in the field test in turn provide an indicator of the percentage of O&M savings available from the general building population. One could argue that this percentage should be verified in a larger sample, but 18.7% O&M savings is sufficient to justify quick adoption of Diagnostic Benchmarking, without waiting for refinement.

Universal benchmarking in New York City (NYC 2012,5) cites a potential for 18-31% O&M savings. This suggests our measured O&M savings (18.7%) are roughly representative, or perhaps lower than in New York's general building stock.

What Changes to Business As Usual Are Needed?

Based on process observations, the author believes business as usual (BAU) consists of ineffective O&M management through poor measurement. Good M&V generates additional savings, and is not a wasted overhead cost. Better measurements and supporting infrastructure must be built into the process, to empower management. Diagnostic Benchmarking should be integral to finding, capturing, and keeping energy savings.

Professionals in facilities design, and staff involved in facilities O&M, should become more familiar with efficient seasonal energy use ***patterns*** - and how to measure them - in buildings. In the field trials, diagnostic benchmarks were used for screening and commissioning, to diagnose excessive energy use, for equipment right-sizing, and for savings retention – all based on measuring and interpreting these patterns.

For screening we compared buildings to peers, each other, conceptual models, or to base-wide averages, to quickly see which buildings used more energy. Commissioning compared actual operating data to simulated operation (LEED modeling) or to exemplary buildings.

To diagnose excessive energy use, we looked for inefficient use patterns for individual fuels. For example, why should a building run its heating boiler a lot during hot weather? Why should the meter reader simultaneously slip on ice and hear the chiller running? Why does a building use twice as much energy as the one next door with the same age and occupancy? Why does a building's reasonable gas use suddenly double? Why is a building too hot in critical zones when its conservatively sized chillers are running flat out? Why doesn't a new building, with all the "right stuff", perform as expected? How can one apply what one just discovered to the next building? Answering questions like these, as we did, yielded much of our savings.

Successfully posing and answering such questions requires suitable metrics – diagnostic benchmarks. Providing infrastructure to get meter readings and convert them into easily understood graphs is a step forward. Providing both the graphs (timely, with minimal effort by the end user) and the understanding to use the graphs, is the way forward.

Benchmarking all buildings first, then capturing O&M savings in the inefficient ones, allows deferring audits in the most efficient buildings, reducing less productive effort. Down-sizing loads by improving O&M will make some retrofit energy conservation measures (ECMs) look less attractive or have longer paybacks. Right-sizing equipment will make equipment efficiency upgrades less costly and more practical. Procedurally, BAU often requires cost/benefit estimates for all measures, including O&M, to prioritize measures. But estimating O&M savings can take more effort than implementing O&M improvements; it's better to just proceed with no-cost O&M improvements, and learn from measured savings.

Background

Several recent works establish a strong basis for going ahead with Diagnostic Benchmarking. First, "universal benchmarking" was undertaken by New York City (NYC 2012). Findings from the majority of large commercial and multi-family buildings there include: (1) large variations in energy use between otherwise similar buildings; and (2) sometimes (but not always) newer office buildings, and larger ones, have relatively higher EUIs. The NYC report's executive summary (NYC 2012, 5) estimates that savings of between 18 - 31% should be available, much of which could be "achieved very cost-effectively through improved O&M".

Second, analysis of benchmarking results for schools in the Northeast, using EPA Energy Star ratings, found a significant portion of reported schools had low Energy Star ratings (probably implying high EUI), compared to their peers of similar vintage (Vadney 2012, Figure 1). If built to similar standards, this implies faulty design, construction, commissioning, or O&M, for low-rated schools. But the Energy Star ratings didn't reveal causes of these differences.

Concurrently, an impromptu field test of Diagnostic Benchmarking took place during three years (late 2008 through late 2011) at a military base (Lambert 2012b). The base was transitioning from master metering to individual automated metering and reporting ("AMR") of larger buildings (>35,000 SF). Discretionary funds weren't available for manpower, parts, or materials for O&M improvements, so O&M improvements in the field trial were, by definition, no-cost. The test was "impromptu" since it was not a dedicated impact evaluation of Diagnostic

Benchmarking. Compromises took place due to the overall mission, to save energy by behavior modifications and retrofits ***as quickly as possible***. Methods and details of the field test have previously been described (Lambert 2012a, 2012b). Results (aggregated and for individual buildings) are addressed here.

In a dedicated evaluation study, each change identified would have first been base-lined, then accomplished and systematically measured. Instead, O&M changes with savings were implemented as soon as identified and agreed to by the base heating, ventilation, and air conditioning (HVAC) and/or energy management and control (EMCS) departments. Also, some changes were not implemented in time for inclusion in this study, due to late discovery, or to perceived staffing limitations. Failure conditions were not allowed to run; they were corrected as soon as practical. So they produced a sequence of two step changes in energy use (an unplanned spike, followed by a deliberate down-tick), without a complete baseline for failed operation. In other cases, a succession of changes occurred, without a complete baseline between changes.

Despite these few departures from a rigorous impact evaluation, the field test results are quite accurate, and establish a "lower bound" for O&M savings potential in these buildings. This lower bound was ***measured*** using Diagnostic Benchmarking.

Measuring Energy Use Changes

The best-case situation was when there was operating history for both a full baseline and a full post-treatment period. "Full" means weekly average outside temperatures covering the full range of average monthly temperatures found in site weather data. Getting a history for one operating condition typically requires trending for 15 to 30 weeks depending on starting time and weather variability. Then we applied the following equation twice, once for baseline and once for post-treatment data, for an average weather year:

$$E_{\mathbf{annual}} = \text{area} * (\sum \dot{E}_{\mathbf{i}}\,(T_{\mathbf{i}}) * \text{Hours}_{\mathbf{i}})$$

Summation is over 12 months; hourly energy use rate, $\dot{E}_{\mathbf{i}}$, is obtained from using a curve fit equation from weekly energy use versus weekly $\mathbf{T_{out}}$, at monthly average temperature $T_{\mathbf{i}}$.

O&M savings were determined for most buildings by subtracting final measured annualized energy use from either "as-found" annualized energy use or (in several cases) annualized "spike" conditions (see Results). For a few buildings, data didn't support such measurements, but substantial accomplishments in these are summarized in "Other Results".

The Buildings

Tested buildings represent many building types. However, single-family housing, schools (tested elsewhere, Lambert Engineering 1991, Lambert 1998), heavy industrial, medical, and agricultural buildings are not represented. Base single-family housing was being privatized, so little work was done there. The base hospital was operated and maintained by a different agency. The industrial shops reported serve as proxies for many light industrial activities found in cities.

The 31 buildings reported include the largest on a 5,300,000 SF base in Southwest Idaho, and some smaller ones. Building age, size, and occupancy type are summarized in the tabular data. Mechanical systems for non-industrial facilities consisted of central gas-fired boilers and central dry chillers, with piped hot and cold water serving coils in zone air handlers. Newer air handlers were variable air volume (VAV) using reheat. Industrial shops had evaporative cooling and overhead gas-fired radiant heat, and heated and cooled to different set-points than non-industrial occupancies. A few shops had under-floor radiant heat. Many shops had in-ground snowmelt systems to prevent freeze-up of rolling doors. Over half the larger buildings had some building automation. Older shops typically used local control.

Some buildings had been significantly remodeled as noted in Table 1 (age highlighted gray). The base had been converted from a central coal-fired steam system. Later, many older boilers were replaced with condensing boilers. Fluorescent lighting had been converted from T12 to T8 lamps with electronic ballasts. Shops used a mix of fluorescent, metal halide and high-pressure sodium lighting.

Results

Table 1 summarizes building characteristics, energy use and O&M savings by fuel type, and aggregated savings and initial energy use for 25 metered buildings. A full baseline during occupied conditions was deemed required for rigorous measurements. The data in Table 1 form the main basis for the following trending and discussion. Six other metered buildings, with less than a full baseline – but significant findings - are summarized under "Other Results".

Table 1 also shows brief descriptions of the sorts of measures accomplished in these buildings (NOTES column), and the number of retrofits (ECMs) identified. Other highlighted entries are explained under "Uneven savings" and "Savings from correcting unexpected faults". Space permits only an overview of results, see (Lambert 2012b) for additional details. Example data and savings calculations, and trend plots of Table 1 data, will be shown at the conference. The data presented here are especially noteworthy because they are the first accurately measured building-by-building O&M savings results with a sample large enough to analyze for trends.

Measured O&M Savings

Energy savings averaged 18.7%, as a percentage of total "as found" EUI for all 25 buildings. This represents only completed work, and would have been substantially higher, had all O&M and retrofit measures been completed. Several trends are evident in the data.

Uneven savings. Savings percentages were uneven; ten "dry holes" (lower-priority buildings) out of the 25 buildings had not yet shown any O&M savings (they might have, had the trial continued). Two implications are: First, if we had excluded buildings without savings, average O&M savings percentages would be larger – for gas, 38% and for electricity, 8%. But generalization to the overall building population must include buildings with no savings.

Second, instead of a steady savings yield, we see "dry holes" (highlighted yellow) and "jackpots". Some buildings yielded large savings and others had warranted less attention (due to lower EUIs), or weren't accessible. We may have missed some small savings opportunities, but achieved very substantial average savings, by prioritizing work in buildings with high EUI. The uneven savings results make effective methods for triage, prioritizing work, and ways to prospect for savings especially important.

Savings from correcting unexpected faults. Three of the largest savings resulted from detecting energy spikes and fixing unexpected faults that did not affect comfort. These failures involved parameters that were not trended by the associated building automation system (BAS).

Without benchmarking, these could have persisted, undetected, indefinitely. All three cases (highlighted blue) occurred *before* tuning, in newer more complex buildings with automated control.

Table 1. O&M savings by building

BLDG #	BLDG AREA kSF	OCCUPANCY GROUP and Type	BASE-LINE GAS mmBTU / year	GAS SAVINGS mmBTU / yr	GAS SAV %	BASE-LINE ELECT mmBTU / year	ELEC SAV-INGS mmBTU / yr	ELEC SAV %	INIT-IAL EUI kBTU /SF-YR	AGE	ECMs	ACTIVITY TYPE, NOTES
		OFFICES										
196	27.52	Off/Warehouse	2387.1	2063.5	86.4	2133.6	425.0	19.9	164.3	11	1	1,4,7
261	17.80	Office	1149.5	710.3	61.8	1484.7	145.5	9.8	148.0	4	0	4,8
510	8.35	Communications	59.7	0.0	0.0	803.6	0.0	0.0	103.3	?	0	11,no access
512	93.38	Office	3544.9	731.0	20.6	6141.8	213.2	3.5	103.7	22	1	1, 2, 7
920	41.72	Electronics shop	3220.1	1304.9	40.5	4714.1	711.3	15.1	190.2	11	1	2,4,7,11
2215	61.47	Office	3762.1	0.0	0.0	4001.8	0.0	0.0	126.3	23	0	11,no access
		INDUSTRIAL SHOPS										
198	43.31	Aircraft svc/rep	2952.5	0.0	0.0	1993.1	0.0	0.0	114.2	12	0	2
200	35.33	Aircraft svc/rep	9123.9	3665.2	40.2	1524.1	63.3	4.2	301.4	13	1	4,5,7,11, 13
211	50.95	Aircraft svc/rep	9298.1	1788.7	19.2	594.3	0.0	0.0	194.2	68	2	5,10
1224	70.36	Aircraft svc/rep	4039.4	737.6	18.3	2206.7	0.0	0.0	88.8	57	0	5,6
1225	46.00	Aircraft svc/rep	3288.6	0.0	0.0	927.9	0.0	0.0	91.7	38	0	10
1330	42.20	Aircraft svc/rep	15729.6	7031.9	44.7	3247.9	621.7	19.1	449.7	56	0	6, 13
1333	35.65	Aircraft svc/rep	4750.1	0.0	0.0	2743.8	0.0	0.0	210.2	56	0	10
1339	43.75	Aircraft svc/rep	2253.9	0.0	0.0	1421.1	0.0	0.0	84.0	16	0	
		MULTI-UNIT HOUSING										
2312	41.62	Dormitory	1603.6	97.0	6.0	1605.5	72.3	4.5	77.1	15	0	12, resets
2314	55.71	Dormitory	2357.6	0.0	0.0	3362.2	0.0	0.0	102.7	12	1	7,9
2320	39.52	VOQ (Hotel)	635.2	0.0	0.0	1401.9	0.0	0.0	51.5	23	0	3
2412	49.08	Dormitory	2954.3	2594.1	87.8	3197.7	362.5	11.3	125.4	13	1	7,8
2416	43.76	Dormitory	1476.0	0.0	0.0	1176.0	0.0	0.0	60.6	16	0	
8395	51.88	Temp. Housing	1612.8	0.0	0.0	1558.0	0.0	0.0	61.1	1	0	
		OTHER										
195	39.32	Club-dining/rec.	2634.5	1979.6	75.1	2111.5	19.9	0.9	120.7	34	0	3,8
206	34.97	Fire station	2588.6	194.3	7.5	2253.0	0.0	0.0	138.5	5	0	5
2371	68.67	Fitness Center	9189.2	2142.2	23.3	5410.3	0.0	0.0	212.6	51	1	12
2700	84.58	Big Box Retail	2792.6	2475.2	88.6	6396.6	116.0	1.8	108.6	23	1	4, 8
2706	53.53	Grocery	1435.6	290.0	20.2	6165.4	0.0	0.0	142.0	15	0	12, 13
TOT	1180		**94840**	**27805**		**68577**	2750.6					
AVG	44.1		3793.6	1288.1	29.3	2743.1	155.7	4.0				
FOR ALL FUELS			**163416**	**30556**	**18.7**	<< includes "dry holes"						
NOTES (1) Significant O&M accomplished before baseline was established, these savings not shown												
	(2) Some requested O&M changes not yet implemented; pending savings not shown.											
	(3) Post-remodel Commissioning was a major activity											
	(4) Re-commissioning (after mission change) or retro-commissioning was a major activity											
	(5) Curtail underfloor radiant heating use, summer-time heating use, and/or doors left open											

	(6) Portions or all of building were ineligible for retro-fit funding			COLOR CODES:
	(7) Fans or pumps need Variable Frequency Drives			DRY HOLES
	(8) Energy Spike; Repair or re-commission failed equipment & control			ENERGY SPIKE
	(9) Avoided equip. damage; no savings shown			PREVIOUSLY REMODELED
	(10) Needs insulation	(11) Needs Economizer	(12) Misc Tune-up	(13) "Hawthorne Effect"

A fourth instance (a coil freeze-up) did impact comfort, but was discovered in time for shut-down of similar equipment, avoiding expensive damage. Avoided damage costs are not shown in Table 1, but the avoided energy costs were 4.6%, nearly 25% of measured savings. Such savings suggest continuous diagnostic benchmarking should be considered for all buildings, even if they have been modernized and properly tuned. Newer more complex buildings may be especially vulnerable to equipment failure and human error.

Savings by fuel type. There were at least eight retrofits identified that promised substantial electrical savings, but most of the measured O&M savings (91%) were of gas. Most unneeded gas uses didn't adversely impact comfort, so they were "invisible." Identifying such uses, and diagnostics to detect them, also made effective methods for triage, prioritizing work, and strategies for prospecting for savings especially important.

Other Results

Six buildings (271, 1327, 2318, 2422, 2610, 2805) lacked complete data for inclusion in Table 1. But there were significant accomplishments in each one.

Building 271 had neither an electric meter, nor a complete baseline, but useful results. Gas data showed 19 % savings after reducing excess outside air (OSA). BAS data suggested OSA leaking into a plenum return, a likely problem in many newer buildings on base.

Buildings 1327, 2318, and 2610 were benchmarked for post-remodel commissioning. Early data showed good performance in 1327 and 2318, compared to peers. Building 2610 had problems, with large reported shortfalls from model-predicted performance (Lambert 2012b).

Building 2422, a dormitory, was metered briefly before a remodel. Data supported right-sizing new boilers to half the original capacity, with substantial equipment savings. This was determined using the limit of heating energy use rate as weekly coincident outside temperature approached the site design temperature. ***Most buildings*** will need boilers or other replacement equipment during their useful life. Original equipment often is oversized, or becomes oversized after efficiency improvements. Right-sizing equipment at the time of replacement offers substantial savings that can offset the costs of installing metering equipment and benchmarking.

Building 2805, a bowling center, had just been remodeled. Scatter in energy use showed erratic operation, traced to dueling thermostats that were causing overlapping heating and cooling. Sequential corrections precluded timely measurement of savings.

Prospecting for Savings

Uneven savings, between buildings and by fuel type, imply several questions. Do readily available building characteristics indicate potential savings? Which buildings should be screened with Diagnostic Benchmarking? What are good savings prospecting strategies?

Possible Indicators of Savings Potential

Annual EUI as an indicator. All else being equal, higher EUI probably indicates more O&M savings potential. In earlier work (Lambert Engineering 1991, Lambert 1998), identical school buildings showed large measured differences in energy use that could only be due to O&M differences. So measured O&M savings were trended versus as-found EUIs.

For all buildings, and within occupancies, higher EUI suggested higher measured O&M savings. But the correlation coefficients ("R^2") were generally low (< 0.65), except for the more homogeneous industrial "shops" ($R^2 \sim 0.95$). Some buildings showed O&M savings, while others were candidates for retrofits. So high EUI, with or without peers, predicted higher O&M savings potential, but not precisely, due to non-homogeneity.

This resembles what EPA's ESPM does with combined energy use. Otherwise similar buildings with low annual EUI get an "Energy Star" rating, while the rest presumably show potential for improvement. But, if high EUI (low ESPM rating) shows something is needed, should one pursue O&M savings or retrofits? A common response to this question is to recommend energy audits; some universal benchmarking jurisdictions even mandate audits (c.f. NYC 2012). A likely assumption, for those who do not recognize that improved O&M can yield large savings, is that audits and retrofits, and large expenditures for them, are in order. This assumption can lead to several unfavorable outcomes:

- Spend significant funds for audits that focus mostly on capital upgrades.
- Overlook large O&M savings potentials and significant interactive effects.
- Scramble to fund and execute inappropriate or over-sized capital upgrades.
- If capital is lacking, find a way to delay or do nothing.

Diagnostic benchmarking provides more balanced outcomes. It often facilitates better choices as to which fuel should be saved, and whether heating, cooling, or lighting should be the focus of efforts. Sometimes, only O&M improvement, or retro-commissioning, is needed. With a more detailed understanding of how the building uses energy, audits, if still needed, can focus on investigating specific issues revealed by diagnostic benchmarks. Capital upgrades, if still required, can be more realistically sized, lowering financing risk and capital outlays. Finally, contracted efforts for O&M or capital upgrades can be better managed, using diagnostic benchmarks for quality control or acceptance of work. Diagnostic Benchmarking provides much better insights into such issues, enabling more effective savings prospecting and management.

Building size as an indicator. A linear regression of percent measured savings versus building area, for buildings with savings, shows that total savings as a percentage of total EUI trended lower with increasing building size. This could mean that larger buildings got more attention to good O&M practices or commissioning, their more prevalent BAS systems affected savings potential, or that O&M staff assigned to smaller buildings were less skillful.

For future savings, several things look worthwhile. First, we should benchmark smaller buildings. New York City now mandates benchmarking for buildings of 50,000 SF and up, but this would have neglected most of our sample (average area = 44.1 kSF), and missed savings. Our sample included buildings as small as 17,800 SF with good O&M savings. Second, increasing percentage savings potential in smaller buildings suggests savings may be constrained by workforce limitations, with only larger buildings getting more "profitable" levels of attention.

We should be looking at increased staffing, improved training and coordination, or an enhanced “toolkit” for O&M personnel - perhaps all three. At the base studied, the HVAC and EMCS shops reported that they were hard-pressed to cover any unplanned workload, beyond regularly scheduled routine maintenance. They sometimes worked at cross-purposes. The author was told that implementation of some requested O&M improvements, not accomplished during the field trials, would require contracting out for additional help.

Job descriptions for HVAC personnel didn’t include responsibility for energy efficiency; they were repair and service mechanics. However, HVAC personnel sometimes asserted “ownership” of building “sequence of operation” (SOO) on grounds they needed such control to meet their responsibility to keep things running. The EMCS shop was theoretically responsible for efficient operation, but didn’t have manpower to do anything other than to maintain the EMCS infrastructure; they were below authorized strength, but constrained from filling openings. Better training, increased staffing, agreements to maintain unified control of SOO, and diagnostic benchmarking as an added tool, would all improve this situation. Such changes would create “green jobs”.

In civilian buildings, staffing, training, and available toolkits are probably a similar constraint. Our sample had few buildings with third-party O&M, too few to claim that it replicated the NYC finding that some larger buildings had high EUI. But owners of civilian buildings built for investment sometimes out-source O&M. Third-party property managers should be given contractual goals and incentives (based on diagnostic benchmarking) and resources to accomplish O&M savings and retrofits.

Building age as an indicator. We expected older buildings would offer more savings potential than newer ones, especially for retrofits. But the highest measured percentages of O&M savings occurred in newer buildings. Some older buildings needed more insulation, but they had simpler systems, and their boilers and lighting had been modernized. Regression of measured O&M savings percentage versus building age was trendless, with lots of scatter. After removing one influential outlier (age 56 in 2010, with recently added energy intensive equipment), newer buildings showed greater savings.

Some newer buildings showed both higher EUI and higher O&M savings. They had “the right stuff” in terms of insulation, good BAS equipment, and relatively more efficient mechanical and lighting systems. But six of them had required re-commissioning or retro-commissioning. Despite the best of intentions, the base may have recently built some “retro-commissioning opportunities”. Alternately, if newer, more complex buildings are inherently more prone to becoming O&M problems, continuous diagnostic benchmarking, to catch problems as they crop up, would be especially important for them. Either way, age by itself wasn’t a good indicator of O&M savings potential.

But a *combination* of high EUI and low age was a particularly good indicator of high O&M savings potential. New York City data may also reflect this tendency, as does data in figure 1 of (Vadney 2012). Vadney’s scatterplot shows roughly 25% of buildings in the upper left-hand quadrant (more recent vintage, lower Energy Star rating). Vadney describes the Energy Star ratings as uncorrelated with building age.

Selecting Buildings for Screening

The data suggest most existing buildings should be screened using diagnostic benchmarking. Buildings down to 15,000 SF, certain smaller high EUI occupancies such as food preparation and laundries, and buildings pending equipment replacements or remodels should be screened. Also, when seeking no-cost savings, lack of capital budget or limited remaining building life should not preclude screening. However, existing metering may be a constraint.

Strategies for O&M savings with Diagnostic Benchmarking

Observed uneven savings, and predominance of gas savings make strategies for prioritizing, finding and capturing savings important. Such strategies have two parts. First, provide procedures for local energy managers. Second, provide policy, support, infrastructure, tools, and training, to empower local energy managers to work effectively.

Local Activities

Whether you're responsible for O&M in one building or a whole complex, the first priority is to continuously diagnostic benchmark all buildings over about 15,000 square feet. Immediate savings or not, you'll want benchmarking to tell you about inevitable failures, so they can be fixed promptly, minimizing waste. Benchmarking will also provide information for right-sizing equipment, and commissioning, when the time comes for replacement or upgrades.

Read the meters yourself (as we first did), if you have no AMR. Work buildings with high EUI first, especially newer ones. With multiple buildings, you don't need a year's data to see which ones use more energy compared to their peers. But with just one building, you'll need context. Context could come from examining similar buildings in an upgraded version of ESPM.

The rest of this to-do list depends partly on site-specific details. Writing a complete tactics manual is beyond the scope of this paper, except for tactics (discussed later) relating to finding gas savings. But infrastructure is needed to execute the to-do list.

Infrastructure Needs

Persuading front-line energy managers to try new methods should be as painless and barrier-free as possible. But some energy managers won't have the metering they need, nor the skill, time, and determination to read, process, and understand data from those meters by themselves, so new infrastructure is needed. After they experience success with diagnostic benchmarking, many will continue to use it. But first-timers, faced with unfamiliar tasks, may never start. Experiencing self-evident measurable success with diagnostic benchmarking will be the best motivator. The infrastructure should be provided to help first-timers to quickly experience success; first-timers need incentives, not unneeded hurdles.

New metering. Some facilities (e.g. some college campuses) don't have metering for individual buildings. All-electric facilities won't have separation of heating and cooling on different meters. This is important to spot problems like simultaneous cooling and heating. We should find ways to finance or subsidize such additional metering.

Such new metering should be AMR capable (with automated reporting of hourly data). This saves the effort to get "eyeball" meter readings and transform them into weekly data, and provides hourly data that can be used to check for good scheduling practices. This way, instead of another task – meter reading - the local energy manager gets an added information resource. Utilities should provide hourly data from revenue meters to the customer location.

Weather data. Someone must collect, screen, clean, and tabulate hourly weather data for a nearby reporting weather station. Our experience showed there were often gaps in this data. NOAA or EPA should provide such data, complete and un-interrupted, in near-real-time hourly form, for a reasonably nearby station, via the internet.

Merging meter and weather data. Next, data must be put into graph form, so energy managers can review and interpret it. EPA already reliably accomplishes data processing – free of charge - for monthly meter data, to provide ESPM data to users. Given weekly or hourly meter data and weather data in appropriate form, EPA could continue to provide ESPM (in more precise form), with spreadsheets that generate suitable graphs, for local use.

Using Diagnostic Benchmarks to Find Gas Savings

Most of the measured O&M savings (91%) were of natural gas. Because savings were dominated by gas, it's useful to summarize how we identified unneeded gas use, and its causes. Captured gas savings resulted mostly from finding and reducing gas use that was inobvious and usually didn't impair comfort. Some of the findings can be generalized to heating with steam or electricity, if those heating fuels are metered.

Space doesn't permit a lengthy discussion of how to use diagnostic benchmarking here. Numerous examples are available (Lambert Engineering 1991; Lambert 1998, 2012a, 2012b). Benchmarking with weekly or hourly data is "diagnostic" in multiple ways. In this case, we're looking to benchmark using a meter that includes heating energy, but (hopefully) excludes cooling energy. With that benchmark, and ideas about what may be wasting gas or electric heat, we test those ideas against the benchmark, and by changing things and observing what happens.

Finding gas savings opportunities. We found high gas use mainly in two ways. The first triage identified relatively high gas use during heating conditions. The second identified buildings that used gas heating during either moderate or summer temperatures.

Utility history at the base showed steady yearly electrical use reductions, but little change in gas use between 2003 and 2009. There were probably improvements in gas efficiency during this time, but reductions in electric use may have shifted more heating load to gas heating equipment. When 30+ new gas meters came on-line in early 2010, we went to work. A plot of base-wide average gas use per square foot, versus monthly outside temperature had been constructed from monthly bills. As each new gas meter was verified functional, we plotted the first week's gas use (and coincident outside temperatures) on the already-available base-wide plot. Buildings that were significantly above base-wide average seasonal EUI stood out immediately, with one week's data.

Applying "triage", some buildings appeared to just need insulation; they could wait while that was confirmed. But several were so high, they obviously needed priority attention; buildings 1330 and 200 topped that list. Buildings 211 and 2371 were also high. Later, full-season baselines confirmed these four buildings had the highest annual gas EUIs on base. The HVAC shop was already working equipment issues at 2371, which also awaited a solar water heating retrofit. But we had identified three buildings with high gas use, needing further investigation, within two weeks of meter installations. While investigating, we collected baseline data for spring and summer.

We already knew summer gas heating was a problem; base-wide plots showed gas use didn't flatten out until monthly average outside temperatures were above 65 F. Data from new meters identified buildings with significant gas heating use during late spring and early summer, during weeks with average outside temperature around 55 or higher. "Gas heating" was shown by elevated gas use, still decreasing as temperatures increased. We captured gas savings in 12 of 15 such buildings (including four mentioned earlier).

Causes and cures for high gas use. These were more diverse, with reheat and excessive ventilation common, and often found together. But first, we sought to use the most efficient heating means. Simply turning under-floor radiant heating off, especially as weather warmed (and using alternate heating means) helped in two shops and the fire station (198, 200, 206). Their under-floor heating wasn't insulated from the ground.

Reheat. Reheat happens when you cool with air or water, and then reheat some of the air for comfort. If comfort is OK, nobody complains, but invisible waste occurs.

Reheat happens in many ways. Problems were most obvious with buildings where the HVAC shop declined to shut off boilers during the summer, due to "too cold" complaints. In one (196) we confirmed reheat during a summertime boiler maintenance shut-down, observing a huge drop in gas use, accompanied by "too cold" complaints while boilers were down. In another (920), we detected alternating cooling and heating (due to insufficient thermostat deadband) by lowering heating supply water temperature by 40 deg. F. Alternating cooling and heating still happened, but alternated more slowly, and gas use fell by 40%. Reheat can be aggravated by unneeded fan operation.

Excessive ventilation. Refers to either unneeded fan operation, or bringing in unnecessary OSA. We confirmed over-ventilation with OSA (and corrections to it), by carbon dioxide (CO2) spot measurements or monitoring. Excessive ventilation shows up in several ways. At 2700, the base exchange, fan scheduling (of rooftop units set for 10% minimum OSA) that suddenly went to 24/7 caused a gas spike. After learning how to program the building's stand-alone BAS, original schedules were restored. CO2 checks showed there was still gross over-ventilation, so we switched half of the roof-top units to "fan auto" mode. We saved 88% of the gas use at the failed condition, yet remained over-ventilated.

In 1330 and 200, most gas use was for tempering make-up air during process ventilation. The processes were intermittent, and by simply asking for recorded logs of exhaust fan and process runtime, we got substantial reductions in fan runtime (reductions from full shift) in both buildings – big "Hawthorne Effects". In building 200, we also sought funding for sensing, economizer and exhaust fan controls, and variable speed fans to automate what the top sergeant there controlled manually using his well-educated nose.

Building 261 needed rebalancing of VAV zonal cooling airflows (identified by a temperature survey), and repair of a concealed broken economizer damper (diagnosed by low CO2 during full occupancy), to enable installed cooling to satisfy cooling loads and save gas. Also, OSA was leaking into a return air plenum; that, we didn't solve.

Building 211 needed overhead radiant heat turned off in the spring, summer and fall, and while hangar doors were open. Also, doors were to be kept closed whenever possible during winter. Needed roof insulation was still pending.

Excessive fan operation was a problem in several places. First we scheduled fans to run only when ventilation, heating or cooling was needed (exchange and 196). The dorms, 196, and the largest office occupancy (512) had problems with both too much ventilation and reheat. We used seasonal supply air temperature resets to reduce reheat in 512, and sought funding for VFDs to reduce excessive airflows and fan power at night, and for too much airflow in oversized air-to-air heat exchangers (AAHX), in two dorms.

In other cases, worries about plenum ceiling returns caused us to monitor CO2 and find over-ventilation. In two dorms with AAHX ventilation, we checked the drawings, and found fresh air had been specified at 120 CFM per bed. Monitoring CO2 in the return air showed CO2 concentrations stayed below 550 parts per million (ppm), confirming ventilation was excessive.

In office-like environments where reheat was occurring, we convinced the HVAC and EMCS shops to use seasonal temperature resets for supply air and for supply water to reheat coils. We also requested automatic temperature reset schedules based on outside temperature, but the shops indicated that they weren't trained or staffed to do that, even where BAS capabilities did support such strategies. Our gas and electric savings could have been greater, had we gotten automatic temperature resets implemented.

Conclusions

O&M energy savings in individual buildings have been directly and accurately measured for the first time in a sample big enough to observe trends. In doing so, Diagnostic Benchmarking, with its capability to accurately capture seasonal by-fuel EUI patterns, in near-real-time, was essential. It enabled fast benchmarking, commissioning, savings identification, fault recognition, problem diagnosis, monitoring of corrections, savings measurement, and equipment right-sizing. Continuous diagnostic benchmarking enables detection and correction of inevitable equipment faults and human errors, finding opportunities and making savings sustainable. Diagnostic Benchmarking should be considered a gold standard for M&V, and essential for sustaining efficiency from commissioning throughout a building's useful life.

The data demonstrate no-cost measured O&M savings of 18.7%, and other benefits. Such benefits justify immediate introduction of diagnostic benchmarking into the mainstream of building energy efficiency practices, together with infrastructure supporting its widespread use, without further deliberation. The measured 18.7% O&M savings fell short of total possible savings; time or manpower constraints prevented implementation or measurement of many proposed O&M measures. So the 18.7% savings is a lower bound. The O&M savings foreseen in (NYC 2012), and the possible O&M problems suggested by data in (Vadney 2012) indicate large nation-wide savings potential. Policymakers need to be aware of the large savings potential we haven't yet seized.

Table 1 excludes savings from retrofits, cost savings from avoided equipment damage, and the costs that would have been incurred as a result of faulty commissioning, had benchmarking not been used. Cost savings due to right-sizing of replacement equipment are especially important, since they are probably significant in ***any*** building that gets renovated. The energy savings imply reduced run-time of equipment, and extended equipment life. We haven't attempted to account for such savings, which would further enhance benefits to building owners. The retrofits, identified and verified by Diagnostic Benchmarking, probably would have shown large savings. At least 30% total energy savings appear feasible, verifiable, and sustainable. The M&V capability of Diagnostic Benchmarking is now proved, and should enable learning how to further improve retrofit savings. But neither data nor space allow proof of such savings here.

Waiting for independent replication of results presented here is unwarranted. Diagnostic Benchmarking is an improved method that makes more objective and rigorous use of portions of the same data already used for input to ESPM, which despite its limitations, has already gained wide acceptance as a standard means of benchmarking.

Once O&M measures and retrofits were identified, continued benchmarking made energy efficiency in existing buildings both inherently self-verifying, and sustainable.

For new construction and retrofits, diagnostic benchmarking establishes whether expected efficiencies or savings are really achieved, not a realistic expectation otherwise. Advocates of "deep retrofits" as the sole means of upgrading our building stock need diagnostic benchmarking for quality assurance to verify their retrofits are really needed, and work as promised. Use with monthly data will offer a supplement to various residential rating systems.

Our data show the same tendencies, to large variations in EUI between otherwise similar buildings, and toward higher EUIs for some newer buildings, observed in broader samples such as universal benchmarking in New York City. This suggests large O&M differences, and inadequate commissioning, are widespread in the US building stock. They represent a very large resource effectively tapped only by diagnostic benchmarking.

We are asking other nations to reduce their CO2 emissions. In view of the threats posed by global warming and especially by ocean acidification, we should also make sure our own house is in order. Failure to take quick action on no-cost and low-cost strategies to substantially improve and maintain the efficiency of our existing building stock is not an option. In the absence of such measures, how can we ask others to act? If not us, *who*? If not now, *when*?

Recommendations

First, EPA should immediately support weekly data collection, reporting, and analysis in the EPA's Energy Star Portfolio Manager, as an option to monthly energy bills. Especially, any such implementation should quickly put completed graphs, by fuel, and the capability to modify graphs in spreadsheets, in the hands of users at no cost, with minimal effort required by them.

This effort should include funding for EPA to upgrade ESPM, and for EPA or NOAA to provide hourly weather data for nearby stations as input to ESPM or to users.

Utilities should be required to retrofit AMR to existing facilities (initially, > 50,000 SF), and to install AMR for all new construction over 10,000 SF. Utilities should be required to report hourly consumption data from AMR to EPA, or to customers.

For new construction, major remodels, and pending major equipment upgrades, metering or metering revisions should include whole-building meters (or equivalent virtual meters) for each fuel type. For all-electric buildings, separate metering of either heating or cooling energy use (whichever is easier) should be considered.

Second, third-party property managers should be required to use diagnostic benchmarking to measure and improve efficiency in facilities they manage. This process should use incentives for improvement (based on measurements using diagnostic benchmarking) in contracts between property managers and building owners.

Third, we should train building O&M staff in the use of diagnostic benchmarking, and teach its use in trade school and college level courses in energy technology. Improving energy efficiency should be put into job descriptions of facilities staff, and their managers. Achieving energy efficiency should be included in job performance ratings of facilities staff and their managers, to help set their compensation or bonuses.

Fourth, contracts for new construction over 10,000 SF and for third-party retrofits should call for installation of AMR, and for commissioning and verification of promised performance or performance improvements, using diagnostic benchmarking.

Fifth, building codes, and standards such as ASHRAE standards 100 and 202, IPMVP, and ISO 14001 and 50001 should call for using diagnostic benchmarking for quality control and for finding and measuring savings.

Sixth, for residential utility billings, utilities should report customer energy use in terms of BTU/hr-SF for each billing period, and the coincident outside temperature for the same days as are in the billing period, for the nearest weather station to the residence.

Seventh, on-bill financing for accomplishing retrofits should be facilitated for businesses, and mandated as an option for residential properties. Interest rates on such financing should be either subsidized, or payments guaranteed, to keep retrofit interest costs low and eliminate up-front out-of-pocket costs that obstruct conservation achievement.

Eighth, recommendations or mandates requiring energy audits and/or retrofits should be revised so that buildings to be audited must first be benchmarked using diagnostic benchmarking. Also, buildings meeting exemplary levels of performance, and using continuous diagnostic benchmarking, should be exempted from initial or repeat audits and/or mandates for retrofits, so long as benchmarked performance remains exemplary.

References

EPA (Environmental Protection Agency). 2013. Energy Strategies for Buildings and Plants>Facility Owners and Managers. Internet: Environmental Protection Agency. http://www.energystar.gov/index.cfm?c=evaluate_performance.bus_portfoliomanager#tools

Lambert L. 1998. *Energy Signature Analysis: Radar for Energy Managers*. Proceedings of the ACEEE 1998 Summer Study on Energy Efficiency in Buildings, 3:217-228. Washington, DC: American Council for an Energy Efficient Economy.

--- 2012a. *Diagnostic Benchmarks: Revolutionary Benefits of Eliminating Sampling Error*. (summary of unpublished manuscript "Toward Self-Evaluating Programs: Diagnostic Benchmarking for Continuous Commissioning in Commercial and Industrial Buildings"). In Proceedings of the 2012 International Energy Program Evaluation Conference. Rome, Italy: IEPEC. June 2012.

--- 2012b. *Getting Your Money's Worth: Diagnostic Benchmarking for Commissioning.* Proceedings, ACEEE 2012 Summer Study on Energy Efficiency in Buildings. Washington, DC: American Council for an Energy Efficient Economy. (paper with ACEEE-omitted lines restored, and poster, available from author, llambert@bendcable.com)

Lambert Engineering. 1991. Report to Commonwealth Electric. "Plymouth Schools – Report of Monitoring Results", Bend, OR: Lambert Engineering, Inc.

NYC (New York City). 2012. Kerr, L., H. Beber, D. Hope. *New York City Local Law 84 Benchmarking Report, New York City Mayor's Office of Long-Term Planning and Sustainability (OLTPS), August 2012*. p5 Internet: NewYork City. http://www.nyc.gov/html/gbee/downloads/pdf/nyc_ll84_benchmarking_report_2012.pdf

Vadney, M., B. Fox, M. Mosser, J Mezlo, B Berstein. 2012. *Benchmarking K-12 Schools: How the Building Energy Performance System Continues to Track and Compare Energy Data in the Northeast and Mid-Atlantic.* Proceedings, ACEEE 2012 Summer Study on Energy Efficiency in Buildings. Washington, DC: American Council for an Energy Efficient Economy.