Gross Savings Estimation for Appliance Recycling Programs: The Lab Versus In Situ Measurement Imbroglio and Related Issues

John Peterson, Athens Research
Don Dohrmann, ADM and Associates
Taghi Alereza, ADM and Associates
Shahana Samiullah, Southern California Edison
Steven Westberg, Hiner & Partners
John H. Reed, Innovologie, LLC

Abstract

Estimates of gross savings for California appliance recycling programs have relied upon methods relating DOE test and characteristic data from multi-era sample of refrigerators and freezers (R/F) to the program population (Athens Research 1996, 1998; KEMA, 2004). The result is a *reliable* approach that yields lab-based program population unit energy consumption (UEC) estimates for the evaluated program and for plausible scenarios involving change in program focus. And yet, skepticism about the external *validity* of this approach prevailed throughout the same decade: do lab-based regression estimates reflect the actual *in situ* consumption of the R/F? If there are systematic *lab-in situ* differences in the recycling population, is this relationship *contingent* upon other variables? *Related* issues include: extrapolation to full year UEC given error in short term in situ metering; degradation, and the factors *other than age* that *select for* a recycling population that is characterized by performance problems; the major physical determinants of differences between in situ and laboratory tests.

This paper reports on both data development and analysis relating to:

- extension of the lab data regression/population UEC estimation method to the 2004-05 IOU's recycling program,
- incorporation of a small dually metered sample (ADM, 2006) as a basis for investigating and preliminarily establishing the "lab-in situ relationship(s)",
- evidence on key issues: extrapolation from short term metering to full year consumption, the level of performance problems typical of recycling appliance populations, and
- summary analysis on the causal determinants of differences between lab and in situ results.

Introduction

Program Background

In 2004-2005, California Investor Owned Utilities (IOUs): Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas and Electric (SDG&E), offered a Residential Appliance Recycling Program (RARP) to households in their service territories. The program had only recently (2002) "gone statewide" across the three electric IOU's, after a considerable history with SCE, and a single vendor (ARCA). In 2004-5, the program targeted residential customers for removal of inefficient yet operable (cooling), pre-1991, 14 to 27 cubic foot refrigerators and/or freezers. These age and size restrictions were new, replacing a more effective design with no age restrictions and a 10 cu. ft. minimum size.[1] The primary goal of the program was to reduce energy consumption by prevention of

[1] The California IOU programs have since returned to a requiring no minimum age, and a minimum appliance volume of 10 cubic feet.

the inefficient unit's continued use in the "participant" household, or in another dwelling or business within the subject utility. The program accepted a maximum of two appliances from a household, and offered (a) free pickup, (b) a $35 incentive ($50 for SCE freezers beginning summer 2005), and (c) environmentally safe breakdown and recycling of appliances. The program sought to remove otherwise operable, transferable, and inefficient appliances from the "option set" of households (including the source household) that may be in the "low-end" purchase or gift-recipient market for such appliances.
Due to a variety of factors, including:

- simple cohort or consumption-at-manufacture changes,
- a major change in program eligibility requirements to include primary appliances,
- changes in the freezer/refrigerator mix,
- possible minor impacts of methodology shifts, and
- possible minor impacts of program penetration.

average estimated recycled appliance UECs have progressively decreased, as measured based on a combination of laboratory UEC estimate sample data and regression approaches expanding to the program population (Table 1).

Table 1. Recent History of RARP UEC Estimates

Program	Study	Refrigerator	Freezer	Overall
1994	Barakat and Chamberlin(1996)	2276		
1994	B&K (1996), 18% reduction	1866		
SCE, 1996	Athens Research (1998), KEMA (1998)	2148	2058	
2002, Statewide	KEMA (2004)	1946	1662	1695
Statewide, 2004-2005	Current study, provisional	1775	1406	1729
SCE, 2004-2005	Current study, provisional	1776	1415	1695

Table 2 provides the number of removals accomplished per IOU during 2004-2005.

Table 2. Program Appliance Removals: 2004-2005

	PG&E	SCE	SDG&E	Total
Freezers	3328	14879	2763	*20970*
Refrigerators	23508	105456	16584	*145548*
Total	*26836*	*120335*	*19347*	*166518*

In summary, the California IOU appliance recycling story through 2005 has several key, somewhat interconnected threads: a move toward a statewide focus including coordination among IOU's, inclusion of primary appliances such that an increasing proportion of program activity involves intervening in the disposal of a just-replaced or soon-to-be-replaced kitchen appliance, and setting of guidelines on age and size that forego the substantial savings available for poorly performing appliances in the 10-14 cubic foot range or manufactured subsequent to 1990.[2]

[2] Memoranda from both KEMA and Athens Research following up on the KEMA 2004 evaluation of the 2002 statewide RARP make it quite clear that there are considerable losses in potential savings associated with these guidelines.

Overview of Recent California RARP Evaluations, with Respect to Gross Savings

In a report to ARCA evaluating the 1994 SCE turn-in program, Barakat and Chamberlin (1996) examined several monitored data sources, notably including approximately 1,100 lab-metered recycled appliances that were part of the ARCA Monitoring Program, circa 1993-1994. Their overall findings suggested a lab-based full year UEC of 2,276 kWh for refrigerators. Based on an unreferenced E-Source report, Barakat and Chamberlin recommended a reduction by 18%, to 1,866 – this might be considered the "headwaters" of the lab/in situ confusion in California (Barakat and Chamberlin, 1996: 11).

In support of the 1998 evaluation of the 1996 SCE program, and taking into consideration California regulatory criticism of the auspices of the ARCA Monitoring Data base, a carefully stratified sample of 136 additional recycled appliances were DOE-tested at BR Labs of Huntington Beach, California. These data were added to the existing library of heretofore ARCA-sponsored tests In this study, a literature review was produced that made it clear that the jury remained out on whether in-use UEC of the removed appliance was systematically lower, higher or contingently related to the values obtained by the reliable, standardized, but perhaps unrealistic DOE test. XENERGY made use of the full lab-based UEC values from the Athens "sub-study" – 2,148 kWh for refrigerators and 2,058 kWh for freezers (Athens, 1998; XENERGY, 1998).

In 2004, KEMA evaluated the 2002 statewide program. In the study, KEMA also worked with BR Labs to again augment the existing trove of DOE-tested <u>recycled</u> appliances, adding 90 refrigerators and 10 freezers. This allowed KEMA to follow up on the Athens approach, adding/testing terms reflecting "sample cohort" and various interactions. Although we might reasonably differ on the final specifications (most definitely regarding how age by defrost interactions are handled), KEMA's overall approach is a sound and thoroughly thought--through extension of the approach that has been taken since 1996. KEMA did report considering some other options, including (a) making greater use of "at manufacture ratings" maintained in California Energy Commission, WAPTAC, and other sources, (b) developing lab/in situ dual metered sample for the 2002 study, but rejected both on a variety of practical and data-availability-related ground. The KEMA results based on the sample DOE-test regression combined with tracking data (statewide) reveal a clear drop from previous SCE evaluations – to 1,946 kWh and 1,662 kWh for refrigerators and freezers, respectively (KEMA, 2004).

Throughout, parties to the RARP evaluations had maintained an interest in both (a) not sacrificing the reliable and increasing library of lab tests on <u>recycled</u> appliances and the select subpopulation of poorly performing but operable and transferable appliances that these samples represent, (b) developing a methodologically defensible dual metering sample (subject to both lab and in situ metering) that would support systematic investigation and possible adjustment of the reliable estimates that one obtains from DOE testing representative and/or properly post-stratified and weighted tracking samples.

The dual metering project of 2004-2005, was the outcome of this effort and designed for a sample of 200 appliances, stratified by appliance type, configuration, size, primary and secondary status, and utility territory. A total of 202 appliances received short term metering in situ and were DOE-tested by BR Labs. The resulting data set allows a number of issues, including the lab/in situ relationship and the possibility of an adjustment, to be at least preliminarily addressed.

The current project, then, relies upon an ARCA Monitoring Study sample from 1993-1994 (approximately 1143 records), the 1998 addition of 136 DOE-tested sample appliances (SCE-BR Labs), the 2003 addition of 100 DOE-tested appliances (Statewide-KEMA-BR Labs), and the current 202 appliances from 2005 lab/in situ monitoring (Statewide-ADM-BR Labs). <u>Hereafter, as necessary, we will refer to these as "sub-samples 94, 98, 03, and 05" respectively.</u>

Reviews of Literature

We are not supplying a literature review here; however a number of reviews pertinent to the lab/in situ problem have been developed. These include Athens (1998), KEMA (2004), and ADM's review in connection with the 2004-2005 dual metering work (ADM, 2004). In general, the information found here, whether closely linked to California recycling or more generally relevant to older appliances in various parts of the U.S. are inconclusive with respect to the degree to which the DOE test overstates or understates in situ consumption, or the circumstances that affect the relationship. KEMA, for example, concludes in its review (KEMA 2004: 8-1):

> There is no significant trend between lab results and in situ results. Therefore, there is no definitive basis present at this time for making an adjustment to the lab-metered estimates of UEC. The results of these studies point in different directions. Some studies found that lab tests over-predicted actual energy consumption; others were inconclusive. None of the studies reviewed involved conditions similar to those of the statewide RARP.

General Structure of the Current Approach to Gross Savings

In the current evaluation, we take the following general approach (which can be flexibly modified to reflect certain other alternatives worth consideration):

- ☐ <u>Lab model -</u> estimate determinants of lab consumption, over approximately 1583 lab-test results from sub-samples 94, 98, 03, and 05. This follows through on the previous approaches, and is capable of supplying low-variance estimates for not only the program as a whole, but meaningful "pockets" within the 2004-2005 program, relating to appliance characteristics, utility program, and for evaluating/assessing the impact of possible changes in program design.
- ☐ <u>Lab/in situ -</u> investigate relationship and estimate possible adjustments regarding lab tests versus in situ results for the same appliance. The general approach here is to develop evidence re. relationship and/or adjustments through either regression or simple estimation of critical ratios, with a strong emphasis upon determining whether the relationships are <u>contingent</u> upon key variables that may be influenced by program design (e.g., focus on secondary appliances, automatic defrost, large households, hotter climate zones, etc.). Secondary work supporting this effort includes developing defensible <u>extrapolation techniques</u>, moving from the observed in situ monitoring period to a full year estimate (briefly reported in this paper). Importantly, this extrapolation is to (a) full year 2005, (b) full year 2004-2005, and (c) full year TMY (Typical Meteorological Year). Additional secondary work, of critical importance to understanding the <u>reasons for differences between lab results and those observed during the actual hours of in situ metering</u>, involves hourly regression analysis of the in situ metering data (briefly reported below).
- ☐ <u>Part use</u> - In the current conception of the impact evaluation (as opposed to previous work), part use (the proportion of the prior year a given recycled appliance had been used rather than switched off), is an aspect or adjustment to gross savings - apart from calculation of net-to-gross as it had previously been conceptualized. This calculation is an "add-on" providing an adjustment to gross savings (which ought to be disaggregate by appliance type and major characteristics, and ideally would be sensitive to what seasons or months plug-ins occurred), and are addressed in the study but not in this paper.

Key Interests and Foci – Gross Savings Analysis

This evaluation, like other useful evaluations, focuses upon generating gross savings results that are (a) *disaggregate*, (b) *portable* in that they are applicable to program planning scenarios, (c) and based on lab model development that is rigorously developed.

As to the lab/in situ issue, the study is committed to recognizing the wide diversity of circumstances that recycled appliances exit on their way to either the recycling center (or the DOE test lab as an interim stop). Therefore we are interested in exploring not only the additive impacts of appliance and household characteristics in accounting for lab/in situ differences among appliances, but also in the key interactions between lab consumption and these characteristics in explaining in situ consumption. It seems highly unlikely that a single overall factor would be an appropriate adjustment for use in any particular jurisdiction, let alone nationally.

Data Collection and Processing: Issues and Challenges

Dual Metering Data Collection

The sample design included attention to representing utilities, appliance types and configurations, sizes, and ages, as well as geographical dispersion within utility territories. A total of 202 appliances were metered, after recruitment, a household survey and a series of one time only appliance measurements, and *then* transported to BR Labs in Huntington Beach for DOE testing.[3] The household survey included information on household size, conditioned vs. unconditioned location, primary vs. secondary status of the appliance, and household income/educational levels. One time measurements were taken with respect to true rms power, voltage, current, power factor, and food load. Continuous metering consisted of a plug-in *power logger,* recording AC current amperage at five minute intervals. Temperature monitoring at five minute intervals occurred interior to fresh food and freezer cabinets as applicable, as well as monitoring of room temperature. Lighting loggers recorded frequency and duration of door openings. From the monitoring data, kW demand per interval is calculated as a product of monitored amps, the one time volt reading, and a one-time power factor measurement specific to whether or not defrost heating is underway. *One version* of the extrapolation procedure multiplies the average of observed kW readings by 8760 hours (ADM, 2006), while others are discussed later in this paper. As in past evaluations, by either Athens or KEMA, the data are transformed into regression-ready terms (binaries, slope terms, etc.) for use in the UEC estimation.

Issues: sample size and vulnerability to selectivity. While in situ metering is necessary to develop estimates of actual use of appliances, there are a wide variety of circumstances (appliance characteristics, climate, location of appliance, household characteristics, etc.) that must be well represented in order to provide useful estimates for any given evaluation. The problem can be partially mitigated by developing a lab/in situ sample that represents wide variation in characteristics and circumstances, such that a model or set of correction factors can be developed to adjust the reliable, standardized, but essentially always unrepresentative DOE model estimates to cover the variation in characteristics and circumstances. However, a sample of 200 is probably only a start! Furthermore, note that obtaining a sample of recycled appliances for lab metering involves far fewer openings for selectivity bias than when in situ metering is involved. Through use of incentives and good customer relations, ADM minimized the validity threat posed by the introduction of an inconvenient intrusion upon the recycling and/or new appliance purchase process. Finally, note that short term metering does not necessarily capture the range of household activity patterns and internal temperature variation that occurs throughout the year. We describe an extrapolation procedure below that was tailored to solve this problem as well as possible with secondary data.

DOE Lab Testing at BR Labs

A description of the DOE test is available at http://www.eren.doe.gov/buildings/codes_standards/. The procedure followed at BR Labs is outlined in 10 Code of Federal Regulations (CFR), Section 430.23 (a), 2001.

A succinct description offered by Meier and Jansky, although slightly dated, captures the important features of the process (Meier and Jansky, 1993: 705), and we add some information from Lloyd

[3] Operationally, ADM intervened in the logistics of program operation by (a) sampling from within scheduled appliance pickups, and (b) sampling from contacts provided through retailers identifying new appliance purchasers with existing appliances needing disposal.

Harrington (2001) as well. Each of these papers offers a useful description of the important differences between DOE and other major testing procedures, including:

- The test chamber is stabilized at 90F.
- The interpolated result (based on systematically varied test conditions) is extrapolated by 365 days.
- There is no ambient relative humidity specification.
- No door openings.
- The fresh food compartment and freezer compartments are empty.
- Freezer and fresh food compartments are served by three thermocouples
- The test incorporates on/off settings of the anti-condensate heater switch.
- Consumption of the appliance is calculated by interpolation of tests bracketing the standard freezer temperature.
- Harrington points out that energy consumption is interpolated for a freezer temperature of -15C (5F), subject to the fresh food compartment being at less than 7.22C (45F). Otherwise, the key interpolation temperature becomes the fresh food compartment at 7.22C (45F). Somewhat awkwardly, where two controls exist, they must be moved together to develop test points. For stand-alone freezers, the key interpolation point is -17.8C (0F).

The major variables, one might hypothesize, determining differences between a given appliance's consumption over a week long period in the home, and the result from a subsequent DOE test include: average temperature (vs. 90F), opening frequency (vs. none), fresh food load (vs. none), and possibly interactions among these factors.

Issues. The DOE test provides standardized results, useful in providing comparisons among appliances at both birth and death (recycling), and in assessing degradation from birth to death. The test will fail to exactly mirror any one appliance's performance in situ, but can serve as a valuable "anchor" by which to efficiently leverage in situ results toward estimates covering a wide variety of appliance circumstances.

Tracking Data Collection

Although never as easy a process as the evaluator foresees, there were significant difficulties associated with this process. Two of the utilities – SCE and SDG&E – continued to work with ARCA, and the tracking data necessary for gross savings estimation continued to be available. PG&E, however, worked with another implementing vendor, JACO, and, beyond some issues relating to consistency between JACO and PG&E records describing picked-up appliances, fewer tracking system variables were available than had been provided in the ARCA format. Table 3 describes the PG&E situation.

Table 3. Data Available Without Imputation from PG&E Tracking, 2004-2005

Tracking Data Available	Frequency
Type	70
Type, Manuf_yr	52
Type,Size	380
Type,Size,Manuf_yr	26334
Total	26836

All of the PG&E records lacked a configuration and defrost specification, and handful (502) lacked type, size, or manufacture year, from the tracking data. At any rate, to support our need for more data describing the tracking population, we:

- Developed large lookup tables supporting both the JACO/PG&E information supplement, but also addition of at-manufacture data for sub-samples 03 and 05, the latter in order to allow auxiliary analyses relating to degradation, quality of age indicators, etc. These lookup tables were built from CEC, WAPTAC, Kouba-Cavallo sources, as well as the smaller table maintained by JACO,

- Developed probabilistic lookup routines based on model number and ancillary JACO-supplied data, so that we obtained matches of fairly high quality for approximately 50 percent of these PG&E tracking records such that about 40% of PG&E tracking records *now* had complete data on type, size, manufacture year (or year range), configuration, and defrost type, and a further 10% had gained *some* data from the lookup process, with one or more gaps remaining,
- We developed imputation routines that created multiple fractionally-weighted records to fill in, in an unbiased way, the remaining gaps, per appliance, in the five main variables of interest. These were based on quantitative correspondence tables developed from complete data on either (a) ARCA distributions at SDGE/SCE plus the completed lookups for PGE, or (b) the ARCA distributions only,
- In order to obtain amperage estimates for PG&E and data, we developed a pair of regression models, calibrated over ARCA data, for imputation of amperage in the PG&E case,
- We have maintained six separate files to represent PG&E tracking data, in order to develop sensitivity tests. These files represent various combinations of lookup table rigor and imputation strategy. For this paper, we rely upon a PGE population variant that includes loosened matching constraints and an imputation upon ARCA data, and
- In order to support eventual use of lab/in situ models that may be sensitive to either primary/secondary refrigerator distinctions, or conditioned/unconditioned space distinctions, we developed logistic regression approaches to indicate likely primary/secondary, conditioned/unconditioned situations in the tracking data -- imputations that can be used carefully in evaluating sample-based models on population data.

To conclude, the ideal would be to have all utility tracking systems include reliable values on appliance type, age, volume, configuration, defrost type, and rated age. This may be a lot to ask for, but these are of critical value in developing population estimates of program impact, as well as evaluating program design scenarios.

Extrapolation from Short-term Metering to Full year Metering-based UEC

In its dual metering study, ADM entertained a number of methods for extrapolating short term in situ metering results to expected full year UEC's. *A necessarily error-prone method must be invoked to produce a full year UEC expectation.* We refer to only three here:

- A simple 8760 hour normalization of the observed wH/hour obtained from the onsite metering.
- A regression model incorporating hourly consumption and hourly temperature based on monitored data on appliance energy use that SCE and PG&E collected in the early and middle 1990's.[4] This model included intercept terms per appliance to reflect "base load," as well temperature and temperature x month interactions. [Model A.]
- A regression model which incorporates the same hourly temperature and month specifications, but also includes an additive expression of month (so that the hourly temperature x month term truly captures the temperature slope specific to that month). [Model B.]

[4] The PG&E data are the monitoring records analyzed in Dutt et. al (1994), under types "E" and "S," while the SCE data were collected during the 1990's as part of SCE's Residential Appliance Enduse Study (RAEUS), administered by SCE. We carefully associated each of these records with its PG&E or SCE weather station. In the case of the PG&E data, this required some extra "temperature pattern matching" work, because weather station indicators were not provided along with the 1990's hourly temperatures included in the PG&E data set.

The regression models were estimated separately for larger side-by-side appliances, top-freezers, stand-alone freezers, and a set of secondary refrigerators in unconditioned space. They reflect the general observation that appliance consumption varies with outdoor temperatures (mediated by changes in indoor temperature and the indoor-internal cabinet temperatures) – see references as disparate as Meier (1993) and Australian Greenhouse Office (2002), as well as ADM's observation that household activity levels, impacting door openings and food load, are in some part seasonal.

Building on ADM's work very slightly, we modified the two types of hourly regressions, including a single base load term equal to the appliance's mean observed wH/hour. The following is an overview of the regression-based extrapolation structure that we built:

- Estimate models A and B for SS, TF, FZ, and RS (side-side, top freezer, freezers, and unconditioned space refrigerators), using available appliance information in the PG&E and SCE 1990's data,
- Use the regression solution to produce appliance type specific estimates of predicted mean monthly consumption and average annual consumption for each utility weather station, over 2005, 2004-2005, and for each California Climate Zone, under average monthly and annual Typical Meteorological Year (TMY) temperatures. Calculate standard errors for each of these predictions, and
- For each weather station (or climate zone), model type, and appliance category, calculate a month-to-annual-extrapolation ratio based on the above monthly and annual expected values, developing a standard error for this ratio.[5]

With this "lookup table" in hand, which extrapolates and provides at least a notional standard error, one proceeds as follows for any dual metering record:

- Determine the appropriate expansion appliance type (e.g., indoor side-by-side primary refrigerator is probably best served by an SS regression model), the regression model type (we have to date been using type B), and whether an expansion is wanted from the observation period to full year 2005, full year for 2004-2005 , or full year TMY.
- Expand to 2005 full year UEC, same weather station, from, say, March 05 monitoring, using the March-specific 200 expansion ratio for the weather station, and the associated standard error for that ratio.
- Expand to 2004-2005 full year UEC, same weather station, from March 05, by (a) first calculating the ratio between regression predictions for 2004-2005 and 2005 and adjusting the observed in suit consumption accordingly (standard error calculated), and then (b) expanding to full year 2004-2005 using the 2004-2005 March-to-full year ratio.
- Similar to the last step, one can expand from 2005 monitoring to TMY, or to another weather zone or climate zone, by use of step (a) just above, all within the same straightforward lookup or correspondence table.

The extrapolated estimates tailored to specific temperature scenarios may be useful in further analysis, and in program planning scenario development (e.g., planning for activity focused in hotter climate zones or utility weather zones in future years. Table 4 provides an example of extrapolated records from the dual metering data set, for which metering occurred in February-March of 2005. The records are a mix of top freezers, single doors, and upright freezers. One extrapolation provided is the simple annual kWh calculated by ADM by normalizing observed consumption to a full year (EXTRAP 8760). Another is the extrapolation performed, as described above, to expand to full year hourly temperatures averaged over 2004-2005 (probably the appropriate ultimate criterion for evaluating the 2004-2005 programs). A standard error is attached, making the point that all extrapolations from short term to full year

[5] Standard error calculated conservatively, omitting any "discount" owing to correlation of monthly and annual predictions.

are error prone. This standard error takes into account error in developing predicted kWh for mean 2004 February temperatures, mean 2004-2005 February temperature, in ratio adjusting from the former to the latter, and in adjusting from February to full year. It is almost certainly an *understatement* of the error involved, as it relies upon the huge volume of hourly records available to the underlying regression on 1990's SCE/PGE refrigerator monitoring data, which would made more realistic using sample replication methods involving random inclusion/exclusion of the 1990's monitoring records *appliance-wise*, for example, using bootstrap methods.

Table 4: A Handful of Records from the Dual Metering Data Set, Including Regression-Based Extrapolation to Full Year kWh, 2004-2005 Temperature Scenario

ID	CONFIG	INST_DT	REM_DATE	EXTRAP 8760	2004-2005 EXTRAP MODEL B	STD_ERR
RF009	TF	05-02-02	05-02-09	700.00	751.13	1.86
RF010	TF	05-02-02	05-02-09	931.00	999.00	2.47
RF011	TF	05-02-02	05-02-09	456.00	526.13	4.78
RF012	SD	05-02-04	05-02-11	764.00	840.24	2.07
RF013	UF	05-02-04	05-02-11	632.48	691.21	2.46

Short term monitoring in situ necessarily undertakes extrapolation error, of some kind, and it can easily outweigh any perceived benefits from avoiding the perceived biases of laboratory procedures. We have already identified and incorporated error in extrapolation, given, for example the listed standard errors for the extrapolated estimates in Table 4. These should be taken into account more explicitly in the evaluation study than they are in this paper. Additionally, note that, the extrapolation is heavily dependent upon the weather characteristics of the period in which short-term monitoring occurs. As it turns out, the regression-based extrapolations all produce smaller full-year UEC estimates, on average over the full dual metering sample, than the simple 8760-hour extrapolation, because the preponderance of in situ monitoring occurred in warmer months. More than half of the monitoring occurred in five months: May-September 2005, meaning that the regression-based extrapolation would amount to a moderate to considerable down-weighting of the observed consumption.

Data Analysis

A certain amount of analysis has been reported in connection with data development in the previous section. However, this section focuses on:

- a rough description of the amount of degradation occurring in the efficiency of appliances that are recycled, reinforcing old findings characterizing the recycled population.
- an hourly regression analysis by ADM which highlights the impact of key differences between the lab test and in situ environment as determinants of consumption.
- a description of the current evaluation's regression model for producing expected lab test gross impact values for the program population and relevant subpopulations.
- data analysis with respect to lab/in situ relationships, including regression and more simplified tabular analysis.

Degradation

In this section, we do not delve deeply into the determinants of appliance degradation, providing only a simple description of its extent among recycled appliances. In 1996, Barakat and Chamberlin analyzed a variety of data sets on older/recyclable appliances, by "clusters" defined by year of manufacture and appliance characteristics. The results yield up a dismal picture reflecting the birth-to-disposal increase in consumption observed *among recycled appliances:* estimates of consumption increase range from lows of 24 and 33 percent increases (very small cluster sample sizes) to highs of 101-187% (larger sample sizes).

In a follow-up to its main evaluation of the 2002 program, KEMA examined degradation over records from the 1998 and 2003 evaluations that were then matched to CEC and WAPTAC sites in an effort to uncover age-degradation relationship (KEMA, 2004A). Based on a review of graphics in the KEMA report, it appears that the median increase in consumption from manufacturer rating to current test value is at least 1.5 (a 50% increase), and that more than 80 percent of units recycled have doubled their test consumption at time of destruction.

In the following, we revisit degradation as a general issue, relying upon our most reliable matches between the dual metering sample of 2005, and the KEMA-BR labs data of 2003. We obtained our matches from the lookup tables described in the earlier tracking data development section. Over a total of 203 reasonable matches, Table 5 shows the distribution of growth percentages for overall, by year, by appliance type, and by defrost type.

Table 5. Test At-birth to Test at-recycle UEC Growth Percentage, Over Quality Matches between Study Sample Appliances and Lookup Tables

COMPARISON GROUPING	RECORDS MATCHED	MEDIAN PCT GROWTH	MEAN PCT GROWTH	3RD QUARTILE PCT GROWTH
OVERALL	203	41.06	50.37	70.80
2003 STUDY	69	51.71	61.60	76.64
2005 STUDY	134	34.26	44.58	68.79
FREEZERS	10	33.87	57.00	96.15
REFRIGERATORS	193	40.64	50.02	70.74
FROST FREE	190	40.08	49.51	69.21
MANUAL DEF	13	48.27	62.96	88.07

These data are provided to reinforce the notion that recycled appliances come from a subpopulation with significant performance issues, despite their operability.

Results from Hourly Regression Analysis on Dual Metering Sample

The study was addresses the determinants of consumption within the household, among the 202 appliances monitored by ADM, using hourly data, to provide information on the key determinants of in situ consumption. This exercise was also meant to yield up clues toward an understanding of the roles of ambient temperature, cabinet temperature, door opening frequency, and door opening duration in determining the differences in consumption observed within the lab (DOE) and in situ. The analysis approach taken in partially addressing these issues involved estimating a regression accounting for hourly kWh consumption, containing individual intercepts (base load adjustment), monitored cabinet temperature, ambient or room temperature, door openings within the measurement hour, and minutes per door opening (ADM, 2007).

The results for the substantive terms in this cross sectional time series exercise are displayed in Table 6, for top freezer, primary refrigerators.

Table 6. Coefficients for ADM Top-Freezer Cross-sectional Time Series Regression

REGRESSOR	COEFF	STD ERROR	T-VALUE
Cabinet_temp_lag1	0.00331	0.00018	17.9
Room_temp_lag1	0.00335	0.00014	23.71
Door_Openings	0.00467	0.00023	20.23
Minutes_per_door_open	-0.00037	0.00038	-0.98

ADM then used this model to consider the gap between the in situ consumption of the appliances, and the expected consumption were then subjected to the mean temperatures and door openings (0) in the DOE test. ADM's measured average cabinet temperature, room temperature, and door openings were 44.1F, 73.3F, and 0.69 respectively (because minutes per door opening were not statistically significant, ADM did not consider this parameter in the evaluation). By contrast, the average cabinet temperature assumed for the DOE test was roughly 38.1F, based on the average result of the cold setting for cabinet temperature used by BR Labs in the "on" condition for the anti-condensate heater -- this serves to provide a reasonable example, but further work on the gap between in situ and lab conditions ought to seriously consider the cabinet temperatures that are averaged over the DOE test's interpolation. Further, the room temperature assumed for DOE was of course 90F, and door openings were set at 0. Evaluating the regression over significant coefficients only, the consumption differential expected for these appliances amounted to approximately 286 kWh per year, for this specific set of appliances in Table 7.

Table 7. Regression Implications – the Gap Between Lab and In Situ

DIFFERENCE	IMPLIED KWH DIFF
Cabinet temp difference	-176.67
Room temperature difference	490.67
Door opening difference	-28.245
Total difference	285.75

This work is included here to highlight the potential of the dual metering study for understanding the components of the lab/in situ delta. Such analyses can be useful in informing the development of planning/forecasting-relevant studies that include more accessible variables (household size, appliance characteristics, climate zone), and are of course interesting in their own right with respect to understanding the reliable DOE test and its relationship to the highly variable sample data gathered in situ.

Development and Application of a Laboratory UEC Regression Model for the Current Study

The development of the laboratory model in the 2004-2005 study entailed including the three prior samples that have been discussed above. General principles involved in developing this regression include:

- A non-negotiable base set of terms additively representing appliance type, configuration, defrost type, and age. These are necessary not only on substantive terms, but to reflect the various ways, that the samples have been stratified in past years – i.e., by "blocking" the regression on all factors ever relevant to stratification, we prevent confusion arising from the stratification.
- Inclusion of terms reflecting sample year, plus attention to interaction of age with cohort, so that a reasonable attempt to capture age x cohort impacts is included in the analysis.
- Investigation of alternative specifications on age. We eventually settled on ln(age), joining KEMA (2004), based on both explained variance and RMSE-related aspects of "fit."
- Hierarchical development of interactions, always assessing the interaction net of base "additive" terms that define interaction as interaction.
- Maintenance of the criterion for identifying and down-weighting outlier records with extreme influence – the same restriction used in Athens (1998) and KEMA(2004).
- Careful consideration of collinearity diagnostics.

The current version of the laboratory regression model is displayed in Table 8.

Table 8. Regression Accounting for Laboratory (DOE) Annual UEC

VARIABLE	DESCRIPTION	COEFFICIENT	T-VALUE
INTERCEP	Intercept	-422.4106	-0.77
ZFZR	Freezer dummy	169.0536	1.84
ZD_BF	Bottom fzr dummy	595.3794	2.91
ZD_SS	Side by side dummy	-129.3553	-0.34
ZD_SD	Single door dummy	-417.1026	-4.73
ZDFF	Frost free dummy	-445.0348	-1.00
ZAGEL	Age nat log	405.2134	2.15
SIZE	Trkg Cu Ft	43.6478	4.59
AMPS	Label Amps	104.1018	4.83
ZIFF_FZ	Freezer x frost free	319.1097	1.94
ZIFF_BF	Bottom fzr x frost free	-302.0484	-1.28
ZIFF_SS	Side by side x frost free	1451.3206	3.80
ZISS_D	Side-side x amps	-126.4332	-2.88
ZSAMP98	SCE/KEMA/BRLABS sample-1998	-48.9460	-0.69
ZSAMP03	KEMA/BRLABS sample-2003	-435.8978	-5.38
ZSAMP05	ADM/BRLABS dualmtr-2005	-649.2073	-10.30
ZIFF_CL	Frost free x ln(age)	299.8206	2.09
ZAGE15UP	Age 15 up binary	1197.8349	2.61
IA15AGEL	Ln age x age 15 up	-524.9782	-3.08
	Model, error df	18, 1564	
	R-square	0.4337	
	RMSE	751.5023	

Note that the model accounts for appliance type and configuration, defrost type, age, and amperage, before entertaining configuration by defrost type interactions, an interaction between side-by-side configuration and amperage (which has persisted over waves of studies), and sample specific intercept terms. A long standing interaction between age and frost free defrost, encountered by both Athens and KEMA in past years, is also retained. Specifying the age-consumption relationship further, we found that significant improvement in fit resulted, net of all other factors considered, if we specified that age impacts subsequent to age 15 were depressed somewhat.

The model results are applicable to population (tracking) data, for various subpopulations, or for planning scenarios. Table 9 simply provides estimates by appliance type and overall, for the 2004-2005 IOU program taken as a whole. For comparison, we provide expected results using an extension of the KEMA model developed in 2004 for the 2002 evaluation – adding only a term reflecting the new 2005 sample – this is of course an inference based on KEMA's approach to sample membership in its evaluation of the 2002 program (KEMA 2004).

Table 9. Expansion/application of the Laboratory UEC Model to 2004-2005 Population (Tracking Data), Statewide

MODEL		PREDICTION	STD_ERR
ATHENS	REFRIGERATOR	1775	53.37
	FREEZER	1406	82.2
	OVERALL	1729	53.22

EXTENDED	REFRIGERATOR	1775	53.5
KEMA	FREEZER	1366	80.44
	OVERALL	1723	53.28

Laboratory/In Situ Relationship

As we have explained, the 202 dual metering observations from the ADM/BR Labs collaboration have been included in the laboratory UEC model just described. We also have used them to investigate the laboratory/in situ relationship, with a view to determining how relationships between the two kinds of estimates are contingent upon appliance characteristics, climate zone, primary/secondary status, or conditioned/unconditioned status.

We begin by describing a model that was carefully and hierarchically developed to reflect the relationship, taking into consideration, as potential determinants of in situ consumption, a number of variables: the laboratory UEC estimate from BR Labs, appliance type, configuration, defrost type, location in conditioned vs. unconditioned space, the average delta between ambient (room) temperature, household size, and whether the dwelling is located among hotter climate zones. We review the model, and then proceed to consider some of the key interactions involving laboratory UEC values, which were considered and rejected from inclusion – in part due to the small number of cases available to the regression analysis. Table 10 represents this final model, which is case-weighted consistent with the sample stratification plan provided by ADM in its dual metering final report (ADM, 2006), and also is subject to the same moderate-to-severe influential observations restriction that was applied to the laboratory UEC

Table 10. Regression of Extrapolated 2004-2005 "Own-Weather Station" In Situ Consumption upon Laboratory UEC and Key Appliance/Household Characteristics

VARIABLE	DESCRIPTION	COEFFICIENT	T-VALUE
U3_B_YY	Dependent: in situ consumption extrapolated to full year 2004-2005.		
INTERCEP	Intercept	-1546.8790	-3.21
AANNKWH	DOE RESULT - laboratory	1.1072	7.32
ZFZR	Freezer dummy	-100.2853	-0.66
ZUNCOND	Dummy for unconditioned space	-224.3353	-3.01
ZHOTCZ	Dummy for warmer climate zone	144.8669	2.10
ZDFF	Frost free dummy	918.1004	3.42
UZDFF	LABKWH x frost free interaction	-0.5683	-3.54
ZHHSIZ3	Dummy	259.0887	3.78
LOGDELTA	Log(avg room- avg cabinet temp F)	309.1803	2.56
DELTPLUG	Dummy for mean plug on missing delta	-27.2552	-0.15
	Model, error df	9, 190	
	R-Square	0.4938	
	RMSE	463.9250	

The model is based on only 200 records that survived the influence diagnostic screen, and contains some very important effects. All other things being equal, freezers are somewhat lesser in situ consumers (not significant but retained as a non-negotiable base term), and, unexpectedly (but consistent through multiple specifications and checks), use in unconditioned space predicts less full year consumption, while hotter climate zones predict somewhat larger consumption. Frost free appliances tend to have net higher in situ consumption. Laboratory consumption interacts with frost free defrost to strongly discount the lab-in situ consumption relationship. We find that household size (correlated moderately with door openings in the monitoring data set), which we've specified as a dummy indicating size greater than two, is worth a net increase of 259 kWh in situ. Finally, we include the all-important room-to-cabinet temperature delta in the model, along with a trivially important dummy variable that is required to flag the handful of cases where a mean value for this variable was substituted.

We are still considering whether to use this model, a simplified version of this model, or a set of ratios specific to a limited set of appliance characteristics and conditions, as an optional basis for adjusting population estimates based on the lab UEC model described earlier. In order to illustrate the impact of its use, we created a set of hypothetical appliance scenarios, for:

- combinations of appliance type, conditioned/unconditioned space, hot/cooler climate zones, defrost type,
- a fixed average room temperature-cabinet temperature delta, and
- laboratory annual kWh results of 1300, 1500, 1700, 1900, 2100, 2300, 2500, and 2700.

We evaluated the model on these combinations of parameters, and the bulk of outcomes (80%) indicated a reduction, comparing the predicted in situ 2004-2005 UEC to the hypothetical laboratory kWh, with (54%) falling in the 80-100% (of laboratory UEC) range. We provide twenty random examples from the 384 generated scenarios for this model, in Table 11.

Table 11. Scenarios Based on Lab/In Situ Model – a Sample of 20 Units

APPLIANCE STATUS	CONDITIONED?	CLIMATE ZONE	DEFROST	HHHSIZE	LAB UEC	PREDICTED IN SITU 2004-05	PCT OF LAB UEC
FREEZER	COND	COOLER CZ	MANUAL	HHSIZE<3	2100	1981.31	94.35
FREEZER	COND	HOTTER CZ	FROST FR	HHSIZE3+	2100	2109.89	100.47
FREEZER	UNCOND	COOLER CZ	FROST FR	HHSIZE<3	1700	1266.06	74.47
FREEZER	UNCOND	COOLER CZ	FROST FR	HHSIZE3+	2100	1740.69	82.89
FREEZER	UNCOND	COOLER CZ	MANUAL	HHSIZE3+	2500	2458.93	98.36
FREEZER	UNCOND	HOTTER CZ	FROST FR	HHSIZE3+	2100	1885.56	89.79
REFRIG	COND	COOLER CZ	MANUAL	HHSIZE<3	2300	2073.84	90.17
REFRIG	COND	COOLER CZ	MANUAL	HHSIZE<3	2700	2516.71	93.21
REFRIG	COND	HOTTER CZ	FROST FR	HHSIZE3+	1300	1549.91	119.22
REFRIG	COND	HOTTER CZ	MANUAL	HHSIZE<3	1500	1332.97	88.86
REFRIG	COND	HOTTER CZ	MANUAL	HHSIZE<3	2300	2218.71	96.47
REFRIG	UNCOND	COOLER CZ	FROST FR	HHSIZE3+	1500	1288.48	85.90
REFRIG	UNCOND	COOLER CZ	FROST FR	HHSIZE3+	1700	1396.25	82.13
REFRIG	UNCOND	HOTTER CZ	FROST FR	HHSIZE3+	2700	2079.97	77.04
SECOND	COND	HOTTER CZ	FROST FR	HHSIZE<3	2700	2060.83	76.33
SECOND	COND	HOTTER CZ	FROST FR	HHSIZE3+	2300	2104.38	91.49
SECOND	COND	HOTTER CZ	MANUAL	HHSIZE<3	1700	1570.02	92.35
SECOND	UNCOND	COOLER CZ	FROST FR	HHSIZE3+	2100	1627.41	77.50
SECOND	UNCOND	HOTTER CZ	FROST FR	HHSIZE<3	2300	1620.95	70.48
SECOND	UNCOND	HOTTER CZ	MANUAL	HHSIZE3+	2700	2711.94	100.44

Continuing to review the lab/in situ problem, we tested a number of specific interactions in developing the model – anticipating that there would be a certain number of such interactions that rivaled the frost free x in terms of variance accounted for. Interestingly we found very little evidence for this with the exception of a possible interaction with very high room temperature (average room temperature >= 85F x laboratory UEC value), which "competed" quite awkwardly with our handling of the room-cabinet temperature delta. This is not to say that a larger dual metering data set, with more representation of variations in appliance type, age, size, and defrost, would not have uncovered more interactions.

To conclude, without a great deal of statistical fanfare, we review (Table 12) the relationships between lab, simple in situ measurement (extrapolated based on 8760/hours monitored only), extrapolation to 2004-2005 via our temperature based model, and extrapolation to TMY based on the model --- segregating the comparisons by meaningful appliance subgroups. We provide the average laboratory UEC, and then express the average in situ measurements as proportions of that average lab UEC. Table 14, containing simple two-way contrasts, indicates that the overall lab-to-in situ drop off is about 13-15%, but 19% if the extrapolation is to the cooler TMY temperature series. Although there isn't much freezer data, freezers seem to experience a steeper lab-in situ drop off than refrigerators. Very old appliances, for either age-related or placement related reasons, experience a slightly steeper drop off from lab to household than appliances less than 20 years old. Finally, contrary to our personal expectations, the lab/in situ difference is much less for appliances in conditioned than in unconditioned space.

Table 12. Relationship between DOE Lab Result and In Situ Measurements- Dual Metering Sample, Contrasts by Type, Defrost, Age Group, Conditioned Space

	CASES	UEC_LAB	SIMPLE EXTRAP	2004-05 EXTRAP	TMY EXTRAP
CONTRAST: OVERALL					
OVERALL	202	1809.1	0.87	0.85	0.81
CONTRAST: APPL TYPE					
FZR	18	1559.9	0.80	0.81	0.75
REF	184	1833.5	0.88	0.85	0.81
CONTRAST: DEFROST					
FROST FR	177	1829.9	0.88	0.85	0.81
MANUAL DEF	25	1661.5	0.80	0.81	0.76
CONTRAST: AGE GROUP					
AGE GT 20 YRS	89	1908.0	0.85	0.84	0.80
AGE LT 20 YRS	113	1731.1	0.89	0.86	0.81
CONTRAST: SPACE TYPE					
COND SPACE	134	1860.8	0.89	0.87	0.82
UNCOND SPACE	68	1707.1	0.83	0.80	0.77

Our final offering (Table 13) provides contrasts based on three variables: appliance type, defrost type, and space type. Note that with the exception of the small number of freezers, the relationship favoring a tighter lab-in situ connection for conditioned space than for unconditioned space is maintained.

Table 13. Relationship between DOE Lab Result and In Situ Measurements- Dual Metering Sample, Contrasts by Combinations of Type, Defrost, and Conditioned Space

	SIMPLE	2004-05	TMY

TYPE	DEFROST	SPACE	CASES	UEC_LAB	EXTRAP	EXTRAP	EXTRAP
FZR	FROST FR	COND SPACE	1	1043.0	0.89	0.99	0.92
FZR	FROST FR	UNCOND SPACE	2	1066.0	0.93	0.93	0.87
FZR	MANUAL DEF	COND SPACE	4	1359.0	0.64	0.69	0.64
FZR	MANUAL DEF	UNCOND SPACE	11	1769.7	0.83	0.83	0.76
REF	FROST FR	COND SPACE	124	1902.0	0.90	0.87	0.82
REF	FROST FR	UNCOND SPACE	50	1697.5	0.83	0.80	0.78
REF	MANUAL DEF	COND SPACE	5	1405.0	0.88	0.93	0.88
REF	MANUAL DEF	UNCOND SPACE	5	1922.0	0.79	0.76	0.73

Summary and Conclusions

We have described the recent history of a data library of laboratory metering samples, acknowledging and examining empirically some of its weaknesses. We are glad that this data has been preserved and continued to be used, because of the *reliability* that it contributes to the estimation of savings in a program that attracts a particular class of older, operable, and likely transferable appliances. We are also very appreciative of the *dual metering* addition to this data collection effort, which although small allows us to make headway on determining whether and how the reliable DOE test/regression analysis-based results of the past ought to be adjusted, and whether adjustments ought to be contingent upon certain appliance characteristics or conditions. The results obtained so far suggest that a downward adjustment of approximately 10-15 percent seems to pertain overall, but both our regression lab/in situ analysis and the simpler tabular analysis indicate that this is probably not appropriately handled as an across-the-board adjustment. We will push forward to determine the best approach, given this small sample size, recognizing that there is some interaction at work, at least in terms of frost free appliances, and a fairly clear adjustment that can be made in program planning regarding household size.

We have reinforced the idea that analysts have an opportunity, in the form of the accumulated lab sample data set, the new dual metering set, and the generally clean tracking data that is associated with these programs, to continue a reliable UEC-estimation approach that is usefully transferable and capable of dealing with planning scenarios to provide cost effective ongoing evaluation and planning support to California's RARP.

We recommend that all tracking systems include data, from the point of pickup if possible, on appliance type, size, configuration, manufacture year, and amperage. We reported on the work that we did to assemble lookup tables from various sources to overcome problems in one tracking file, and the modeling work necessary to "fill in" certain other parameters useful in scenario testing and other exercises – primary/secondary appliance status and conditioned/unconditioned space. Perhaps the efforts made in this cleanup/fill in modeling will be useful to others.

A set of extrapolation methods was also presented. The regression methods based on early 1990's California monitoring data appear to be fairly robust, and do provide useful, temperature sensitive extrapolations, complete with optimistic standard errors. The general point is that in situ monitoring, in addition to being expensive and difficult to use in adequately representing program populations (and therefore very much assisted by the "leverage" coming from lab data), is *also* based on a sample in time, and is error prone – despite its "validity." In this connection, it would be very helpful to have more long term California-wide metering data, for aged appliances of various types as an adjunct to the data we used. Again, the efforts made thus far may be helpful in other situations involving short term-to-full year extrapolation problems.

In future evaluations of RARP, it would be helpful to have continued inclusion of dual metering approaches. By adding another 200 dually metered appliances, enormous gains can be made in our ability to estimate the lab/in situ relationship for important appliance subgroups and household conditions. In adding to this sample, it is important to seek out variation with disproportionately stratified samples that adequately cover the extremes of appliance characteristics and climate zones, as well as "filling in" for recent lack of small or younger appliances due to temporary guideline changes. By ensuring wide variation in the in lab/in situ sample, its usefulness in challenging, and as appropriate, adjusting the results of the reliable laboratory results, is maximized.

References

ADM, 2004. "Dual Metering Study to Support 2003 EM&V of Statewide Residential Appliance Recycling Program: Research Plan." September, 2004.

ADM, 2006. "Dual Metering Study to Support 2003 EM&V of Statewide Residential Appliance Recycling Program: Final Report." June, 2006.

ADM, 2007. "Comparison of In Situ Extrapolations to BR Estimates." Memorandum, January 8, 2007.

Athens Research, 1998. "Refrigerator/Freezer UEC Estimation, 1996 ARCA/SCE Turn-in Program."

Australian Greenhouse Office, 2002. "Appliance Electricity End-Use: Weather and Climate Sensitivity".

Barakat and Chamberlin, 1996. "Estimating Consumption of Refrigerators in Utility Turn-in Programs."

Dutt, G., J. Proctor, M. Blasnik, and A. Goett, 1994. "Large Scale Residential Refrigerator Field Metering." *In proceedings of the ACEEE 1994 Summer Study on Energy Efficiency.*

Harrington, l., 2001. "A Comparative Assessment of Refrigerator Test Methods." *ECEEE 2001, Paper5.194. Energy Efficient Strategies, Australia.*

KEMA, 2004. "Final Report: Measurement and Evaluation Study of the 2002 Statewide Residential Appliance Recycling Program. February, 2004."

KEMA, 2004. "2003 EM&V RARP Study: Degradation and Market Potential Analysis." December, 2004.

Meier, A, and R. Jansky, 1993. "Field Performance of Residential Refrigerators: A Comparison with the Laboratory Test." *ASHRAE Transactions: 99, pt. 1, pp. 704-713.*

XENERGY, 1998. "Impact Evaluation of the Spare Refrigerator Recycling Program: CEC Study #537."