DataIQ – A Machine Learning Approach to Anomaly Detection for Energy Performance Data Quality and Reliability

Constantine Kontokosta, PhD, PE, New York University, Center for Urban Science and Progress & Tandon School of Engineering
Bartosz Bonczak, New York University, Center for Urban Science and Progress
Marshall Duer-Balkind, District of Columbia, Department of Energy and Environment

ABSTRACT

The paper develops and tests a strategy for evaluating and improving the data quality and analysis of benchmarking data from privately- and publicly-owned buildings. We utilize machine learning tools to support the District of Columbia District Department of the Environment (DDOE) in improving the quality of building energy data collected through its mandatory building rating and disclosure policy. Using energy disclosure data from calendar year 2013, this paper presents a new data quality rating algorithm and grading system – known as the Data Integrity and Quality (*DataIQ*) score – to rank the relative reliability of reported data and provide a tool to improve the identification and prediction of data quality concerns going forward. We apply non-parametric statistical anomaly detection techniques to identify data quality and reliability concerns in self-reported building energy benchmarking data. This approach creates a foundation for more robust and precise analysis of city energy data that can provide policymakers with additional insight to improve the reliability of building benchmarking data analysis and inform the design of data-driven energy efficiency policies.

INTRODUCTION

The District of Columbia and the City of New York were the first two jurisdictions to adopt mandatory benchmarking and disclosure laws (City of New York 2012, Hsu 2014 Kontokosta 2012, 2013). These laws require large buildings (in both cities, buildings over 50,000 gross square feet) to annually benchmark their energy and water performance using the U.S. Environmental Protection Agency's Energy Star Portfolio Manager software, and report the results to the cities, which make the data available online.

Mandatory benchmarking and Disclosure programs have three fundamental purposes: (1) to provide building owners & managers with better information about the efficiency of their own properties, and how those properties compare to local and national peers; (2) to drive market transformation by allowing market actors to easily compare the performance of properties when leasing, buying, or investing; and (3) to provide policy-makers and program administrators with better information for planning, program design, and targeting. (Hart 2015; Keicher et al. 2012).

All of these uses depend fundamentally on the reliability of the reported data (Hsu 2014; Kontokosta 2013; 2015; Palmer and Walls 2015). Building owners, managers, and other market actors must have confidence in the reliability of the benchmarking data, and subsequent analysis, in order to make decisions that save money and energy. And cities, utilities, and contractors must have confidence in the reliability of the data to design policies and programs that address market needs and drive change. If the market comes to question the quality of benchmarking data, the resulting uncertainty will undermine the potential for data-driven market transformation and a precipitate a loss of trust among decision-makers that could be difficult to regain. Thus, both actual or objective and perceived or contextual data reliability are needed to move from the superficial availability of information to proactive decisions based directly on insights derived from it (Wang and Strong 1996). For this reason, improving the completeness and quality of the benchmarking data is a very high priority of many benchmarking programs, including the District's.

In the implementation of the benchmarking law, the District of Columbia Department of Energy and Environment (DOEE), which implements the benchmarking law in Washington, DC, has found common self-reported data errors. In general, we can classify these errors with respect to the completeness of the data, its consistency, and compliance rate (Pipino, Lee, and Wang 2002). Building on the benchmarking data quality issues in Kontokosta (2013), the types of data errors specific to building energy disclosure data can be divided into five categories:

1. **Fatal Errors**: User error that leads to a report being submitted without any metrics, such as Gross Floor Area, Energy Use Intensity (EUI), Water Use Intensity (WUI). These errors are usually caused by the respondent not entering in complete meter data for the whole year or setting the "active date" for meter or space use values incorrectly;
2. **Energy Data**: Inaccurate or incomplete energy consumption information due to either data collection errors (not reporting all the energy meters for the building), data entry error (e.g. incorrect units), or utility company errors (incorrect billing or aggregation);
3. **Floor Area**: Inaccurate square footage based on either the use of unverified tax data square footage, which can often be wrong, or incorrect understanding of what spaces in the building need to be included (e.g. net vs. gross square footage);
4. **Space Use**: Inaccurate space use attributes due to data entry errors, confusion about the requirements, or the use of default values; and
5. **ENERGY STAR Scoring Issues**: Problems in the methodology of ENERGY STAR Portfolio Manager itself (e.g. changes in site-to-source energy ratios, or use of outdated or insufficient reference data sets).

Fatal errors can be minimized when cities work directly with building owners to correct these (relatively) simple user errors. In DC, 25% of initial disclosure reports contained this type of error; however, the District has reduced this rate to just 3% through compliance assistance and enforcement. Starting in late 2015, DC is now enforcing compliance standards for observed fatal errors, considering them to be equivalent to non-submittal. The fifth error category - problems with Energy Star Portfolio Manager - is an important area for academic investigation, but beyond the scope of this paper (for additional information, see Kontokosta 2015 and Hsu 2014). Moreover, the value for city governments of using a federally-supported, industry-standard tool that is common across jurisdictions is an important consideration when evaluating the lack of local control and methodological concerns about the Portfolio Manager software itself.

A fundamental concern of cities trying to implement benchmarking laws are errors relating to EUI and space use values (which drive the Energy Star score). Some jurisdictions, such as Chicago, Illinois and Montgomery County, Maryland, have started to require third-party verification of data quality; in other jurisdictions, the burden to ensure data quality falls more heavily on the city. DOEE does operate a Benchmarking Help Center that fields between 1,500 and 2,000 requests for assistance a year. However, the scale of the data makes a purely manual approach to data quality verification insufficient—DC has approximately 1,600 benchmarking reports, while a larger city like NYC receives over 14,000 each year. Therefore, a methodology to triage reports and flag those with data quality concerns is needed.

The paper develops and tests a transparent and robust strategy for evaluating and improving the data quality and analysis of benchmarking data from privately- and publicly-owned buildings. We utilize machine learning tools to support jurisdictions like the District of Columbia in improving the quality of building energy data collected through its mandatory building rating and disclosure policy. Using energy disclosure data from calendar year 2013, this paper presents a new data quality rating algorithm and grading system – known as the Data Integrity and Quality (*DataIQ*) score – to rank the relative reliability of reported data for the largest building sectors, and provide a tool to improve the identification and prediction of data quality concerns going forward. We apply non-parametric statistical anomaly detection techniques to identify data quality and reliability concerns in self-reported building energy benchmarking data, focusing here on office buildings. The next section describes the data and data cleaning methodology, followed by a discussion of the DataIQ methodology and results. We conclude with a discussion of the implications of the model and its potential application as part of city energy disclosure policies.

DATA AND DESCRIPTIVE STATISTICS

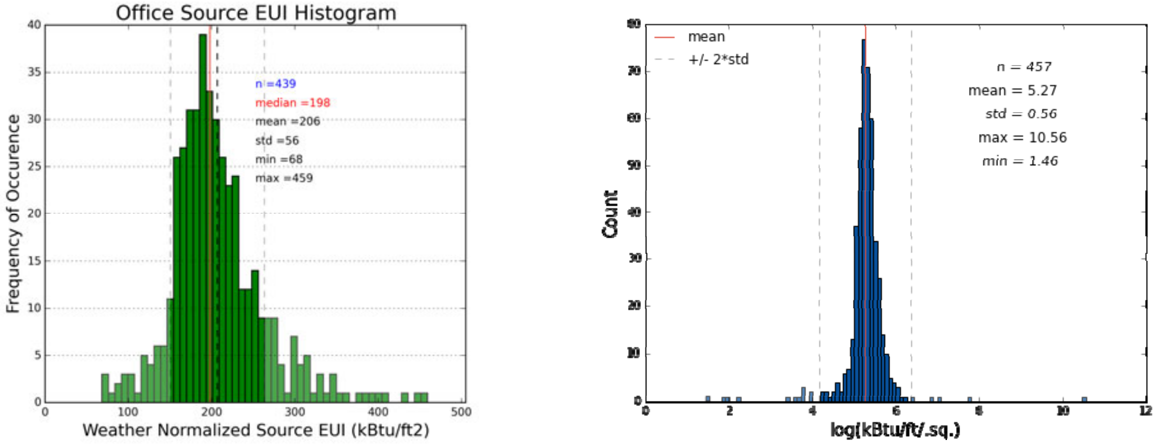
The data for this analysis include all properties subject to the requirements of the Clean and Affordable Energy Act of 2008 that reported data to the Washington, DC Department of Energy and Environment for calendar year 2013. The analyzed dataset consists of all private and public buildings subject to the energy disclosure requirement that provided data through September 4, 2015. After the removal of duplicate entries, the 2013 dataset includes a total of 1,774 unique records, divided between private buildings (1,456), public (government) buildings (274), and public housing (44). Due to non-trivial errors and omissions of various reported fields specific to multi-family housing, we discuss only office building data here.

Multi-stage data cleaning represents a critical component of building energy analytics and the ability to extract reliable insight from energy data. First, the datasets are merged to create an integrated dataset of private and public buildings for each respective year. As part of this process, variable headers and other entries were cleaned of atypical characters and formatting inconsistencies. The combined dataset was then joined with the provided tax data based on Property ID and Square Suffix Lot (SSL) information. This process successfully matched disclosure and tax lot data for 1,169 properties in the 2013 dataset.

Next, entries with duplicate Property IDs were removed, with only the most recent entry retained. Where feasible and appropriate, missing values in the most recent Property ID record were imputed from earlier submissions for the same Property ID. The resulting dataset contained 1,774 for the 2013.

Of the reported fields, Weather Normalized Source Energy Use Intensity (EUI) expressed in kBtu/ft2 is the most widely used energy efficiency metric in energy disclosure analyses (Hinge, Winston, and Stigge 2006; Pérez-Lombard, Ortiz, and Pout 2008). Therefore, the next step in the cleaning procedure removed observations for which an EUI value was either omitted or unable to be calculated due to missing square footage or energy consumption data.

The final cleaning step identified and removed outliers based on the statistical properties of the observed distribution of the selected features. In any self-reported data, data entry errors can constrain the precision and reliability of analysis. In addition, properties that report accurate data, but are significantly different from the rest of the sample, can skew results and lead to the false interpretation of observed trends. Here, we account for outliers by first conducting a log-transform of the data based on EUI, as its unaltered distribution is asymmetrical and has the right-skew characteristic of a logarithmic-normal distribution (fig 1). Taking the natural logarithm of EUI normalizes the distribution and allows for the use of the standard deviation as a threshold to detect outliers (fig 2).



Figures 1 and 2: Distribution of weather normalized source EUI for office buildings (left); Log-transformed distribution of weather normalized source EUI for office buildings (right)

Following the logarithmic transformation, observations greater or less than two standard deviations from the calculated mean are flagged as outliers and dropped from the analysis dataset (fig. 2). This outlier detection methodology was applied by building type, so the distributional analysis is conducted for Office buildings, Multifamily buildings, and “Other” properties independently. The final analysis dataset consists of 1,257 properties in 2013.

Flag type	**2013**			
	Total	**Multifamily**	**Office**	**Other**
Source EUI = 0	315	84	73	158
Source EUI = Null	10	0	1	9
Source EUI final flag (w/outliers)	517	163	92	262
Site EUI = 0	315	84	73	158
Site EUI = Null	11	0	1	10
Site EUI final flag (outlier)	488	145	90	253
Gross Floor Area = 0	80	0	0	80
Gross Floor Area = Null	18	0	0	18
Gross Floor Area final flag (w/outliers)	308	26	60	222
GHG Emissions Intensity = 0	203	53	35	115
GHG Emissions Intensity = 0	70	14	10	46
GHG Emissions Intensity final flag (w/outliers)	462	147	71	244

Table 1: Number of flagged properties by flag type

Figure 3 shows the distribution of properties by year of construction (both office and multifamily properties in the cleaned dataset). A majority of the buildings included in the analysis are relatively new when compared to similar large cities in the United States, such as New York City and Chicago. It is specifically visible among office buildings, with more than 80% constructed after 1960, and several distinct building booms are noted in the 1960s, 1980s, and 2000s. Multifamily buildings exhibited significant growth in the 1960s and during the years between 2000 and 2010, with a majority constructed prior to 1970.

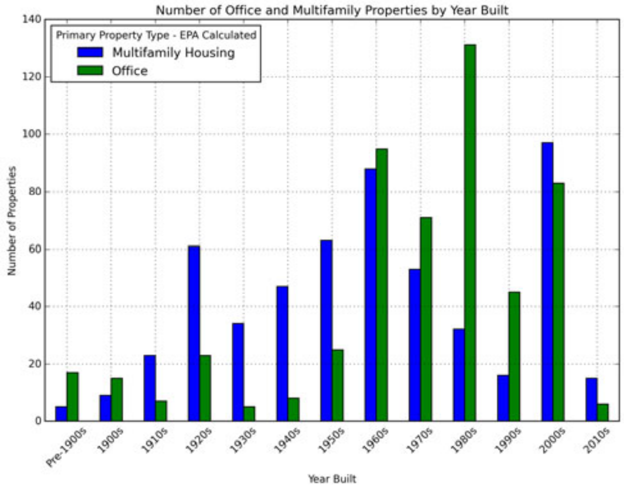
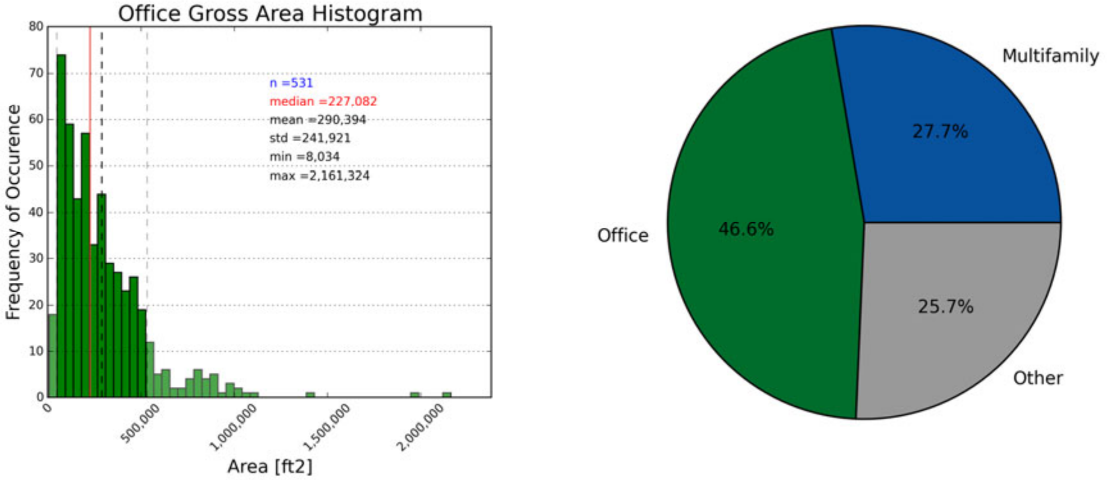


Figure 3: Multifamily and Office buildings by construction year

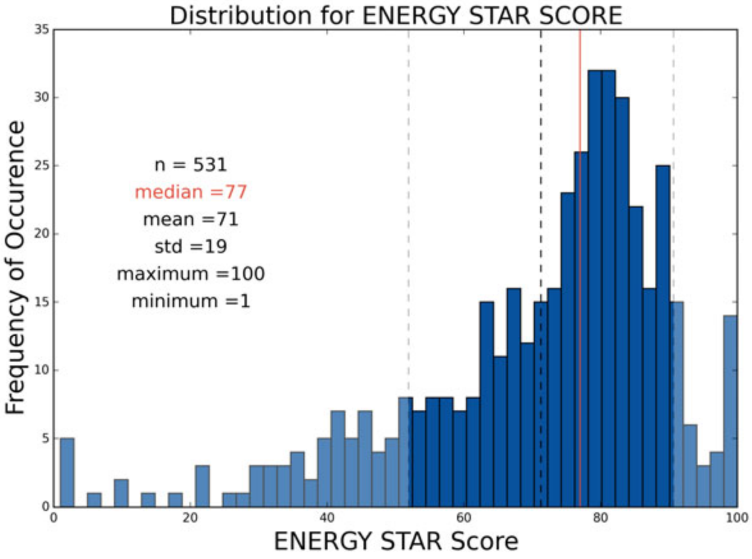
Another important factor energy use is building size. As one can observe in Figure 4, the vast majority of the properties are less than 200,000 ft^2—reflecting, in part, the federal limitations on building heights in Washington, DC. However, office buildings tend to be larger than the multifamily stock, with numerous buildings exceeding 500,000 ft^2 (fig. 4). The total square footage of office buildings accounts for just under half of the total square footage for all buildings included in the dataset (fig. 5).



Figures 4 and 5: Histogram of gross floor area (left); Proportion of each property type by gross floor area in the disclosure sample

The variable of interest in most building energy benchmarking analyses is the Weather Normalized Source Energy Use Intensity (Source EUI), expressed in thousands of British Thermal Units divided by the building square footage (kBtu/ft^2). As shown in Figure 1, office properties have in DC have a median EUI at 198 kBtu/ft^2, as compared to data from the 2003 Commercial Building Energy Consumption Survey (CBECS) that indicates a median source EUI of 210 for office buildings in the Northeast Region.

The U.S. EPA Energy Star score uses a quantitative methodology to assess and rank building energy performance for buildings that elect to submit energy data (Kontokosta 2015). Based on the data received for this analysis, 442 out of 531 office properties received Energy Star scores in the 2013 data. From the 2013 data, the median score for the Washington, DC area was 77, which is higher than national median of 50, and higher than several other cities with energy disclosure policies. The distribution of office building Energy Star scores is shown in Figure 6. This high energy performance may in part reflect the impact of federal policy; the U.S. General Services Administration (GSA) is required to only lease in Energy Star certified buildings unless none are available, and the GSA is the single largest tenant in Washington, DC.



Figures 6: Histogram of Energy Star scores for office buildings

METHODS & RESULTS

We apply non-parametric statistical anomaly detection techniques to identify data quality and reliability concerns in self-reported building energy benchmarking data (Chandola, Banerjee, and Kumar 2009). A multi-variate regression model with robust standard errors is used to measure predicted energy performance, normalizing for a number of factors, including building location, occupancy variables, and other features described below. This allows for a measure of expected energy intensity given a building's attributes, based on the entire Washington, DC benchmarking sample as the reference. A significant deviation between expected and actual energy intensity is an indicator of potential data reliability concerns. The model is given by:

$$y = \alpha + \beta BLDG_i + \gamma OCC_i + \phi AGE + \delta FUEL_i + \varepsilon$$

where BLDG consists of a range of physical building characteristics; OCC includes occupancy variables such as worker density and operating hours; AGE accounts for categorical variables of building age; FUEL represents fuel type mix and the presence of an Energy Star score greater than 75; β, γ, ϕ and δ are vectors of parameters to be estimated; and ε is the error term. Specific variables used in regression are described in Table 2. Feature selection was done using only fields available in the disclosure dataset, namely those available through Portfolio

Manager. This was done to ensure both the ease of interpretation and some level of standardization across cities, so that the model can be more easily replicated and the results compared. Variance Inflation Factors were calculated to identify variables with observed collinearity. A correlation matrix for the included variables in shown in figure 7. For validation purposes, we create an additional field indicating whether there was missing information for any variable used in the model. This allows us to check whether the accuracy of prediction was influenced by incomplete entries, which can yield valuable information about overall data quality.

The dependent variable, *y*, is the natural log of Weather Normalized Source EUI. The model is trained on a randomly selected sample representing 60% of the cleaned dataset based on *Source_EUI_final_flag* (243 observations) and tested on the remaining 40% (162 observations). The flexibility of the model was tested in 100 iterations that showed small variations between observed R-squared values for training and testing subsets (average 0.507 and 0.470 respectively). The model is then fitted against the entire cleaned data set, which consisted of a total 405 complete observations. The actual variable values of the initial, uncleaned sample of 457 office buildings are then multiplied by the respective coefficients values resulting from the model estimation to calculate a predicted EUI. This predicted EUI is then compared to the actual EUI reported for each building to estimate the ratio of predicted to actual EUI. The results of the regression model predictions on training, testing and clean datasets for the 2013 reporting year are plotted in the table 3 and figure 8 below.

Included Variables and Definitions	
Year of Construction	categorical variable (0,1) whether the property was built within given period. Division based on quantile values in order to keep similar subset sizes
Electric Primary	categorical variable (0,1) whether the Electricity ('Electricity Use Grid Purchase and Generated from Onsite Renewable Systems kBtu') covers 50% of total energy usage ('Site Energy Use kBtu')
Gross Floor Area (log)	logarithmic value of 'Property GFA EPA Calculated Buildings and Parking ft2'
Occupancy Ratio	categorical variable based on 'Occupancy' (2 while equal to 100, 1 when less than 100 and equal to 80, 0 while less than 80)
Worker Density	'Office Worker Density Number per 1000 ft2'
Operating Hours (log)	logarithmic value of 'Office Weekly Operating Hours'
% Cooled	categorical variable (0,1) whether the 'Office Percent That Can Be Cooled' covers 95% or more of total energy building area
ES Labeled or equivalent	categorical variable (0,1) whether the 'ENERGY STAR Score' is equal or higher than 75

Table 2: Descriptions of specific variables used in the DataIQ model

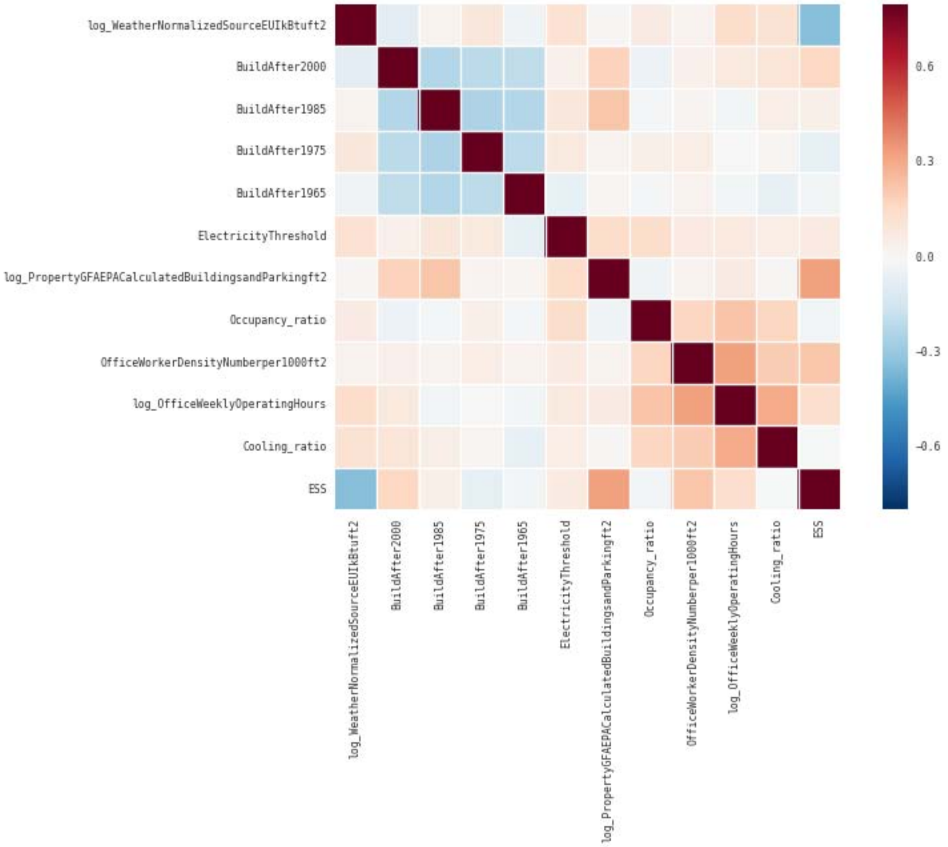


Figure 7: Correlation heat map between included variables

```
      Robust linear Model Regression Results
====================================================================================================
Dep. Variable:     log_WeatherNormalizedSourceEUIkBtuft2   No. Observations:                  405
Model:                                               RLM   Df Residuals:                      393
Method:                                             IRLS   Df Model:                           11
Norm:                                             HuberT
Scale Est.:                                          mad
Cov Type:                                             H1
Date:                                   Thu, 10 Mar 2016
Time:                                           16:19:52
No. Iterations:                                       34
=======================================================================================================================
                                                          coef    std err          z      P>|z|      [95.0% Conf. Int.]
-----------------------------------------------------------------------------------------------------------------------
Intercept                                               1.6035      0.082     19.437      0.000         1.442     1.765
BuildAfter2000                                          0.0083      0.011      0.741      0.458        -0.014     0.030
BuildAfter1985                                          0.0257      0.010      2.452      0.014         0.005     0.046
BuildAfter1975                                          0.0168      0.011      1.596      0.110        -0.004     0.037
BuildAfter1965                                          0.0037      0.011      0.348      0.728        -0.017     0.025
ElectricityThreshold                                    0.0177      0.030      0.584      0.559        -0.042     0.077
log_PropertyGFAEPACalculatedBuildingsandParkingft2      0.0575      0.011      5.217      0.000         0.036     0.079
Occupancy_ratio                                         0.0076      0.007      1.136      0.256        -0.005     0.021
OfficeWorkerDensityNumberper1000ft2                     0.0293      0.004      7.732      0.000         0.022     0.037
log_OfficeWeeklyOperatingHours                          0.1827      0.034      5.409      0.000         0.117     0.249
Cooling_ratio                                           0.0266      0.014      1.857      0.063        -0.001     0.055
ESS                                                    -0.1393      0.007    -18.978      0.000        -0.154    -0.125
=======================================================================================================================
```

Table 3: Results of the predictive model.

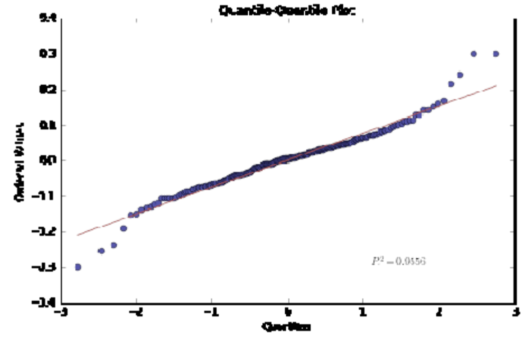
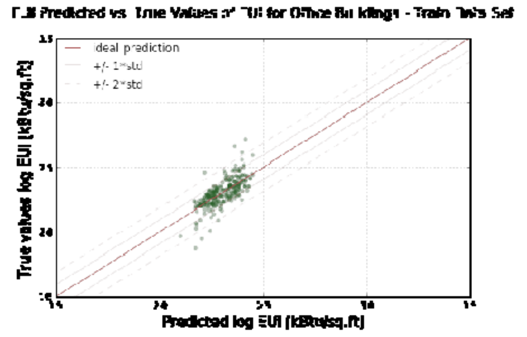
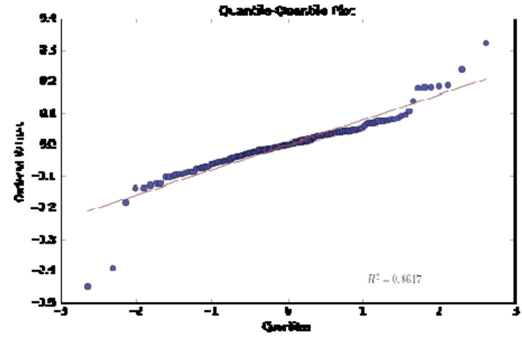
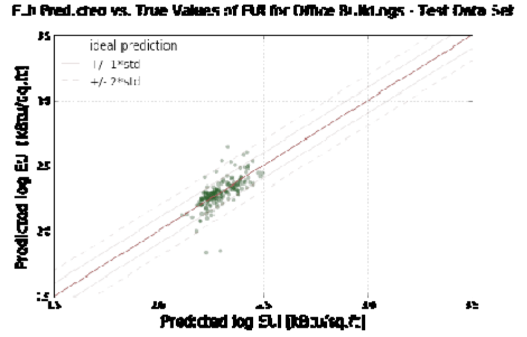
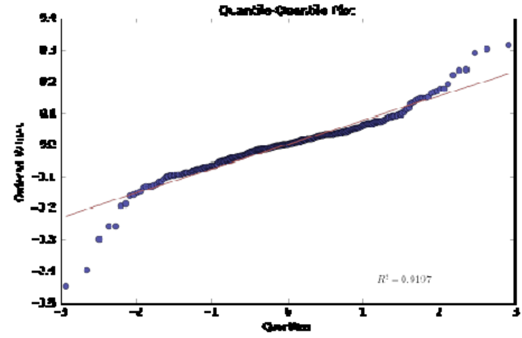
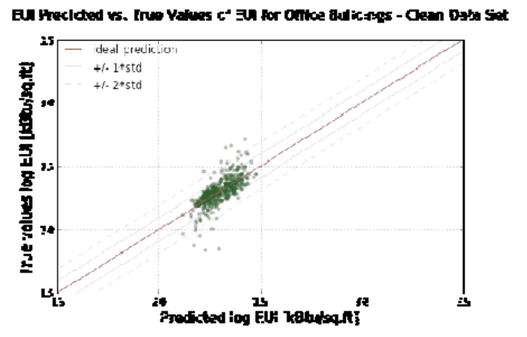
Data set	R2	std	Quantile-Quantile Plot	Scatter plot
Training	0.507*	0.095*		
Testing	0.470*	0.095*		
Clean	0.499	0.094		

Figure 8: Modeling training statistics and results

The final DataIQ scores are determined by the distribution and variance of the ratio of expected to actual EUI, based on the results of the model described above. Histogram represents the distribution of the ratio.

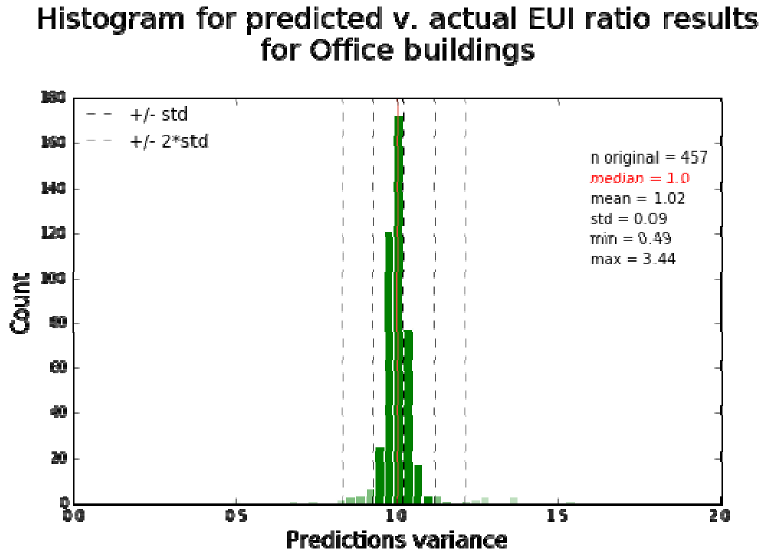


Figure 9: Distribution of the model prediction ratio

A four-point grading system (*A*, *B*, *C*, *D*) is assigned based on the deviation from a "perfect" prediction calculated for the clean data set. The distance is measured by subtracting 1 from predicted versus actual EUI ratio. The grade is assigned with relation to the calculated standard deviation such that:

A - less than 0.5 standard deviations
B - between 0.5 and 1.0 standard deviations
C - between 1.0 and 2.0 standard deviations
D – equal or greater than 2.0 standard deviations

Results closer to zero indicate the relative measured reliability of the data provided for a particular building. The results of the DataIQ scoring based on Weather Normalized Source EUI are presented in Table 4. In addition to graded properties, there are 74 observations without the score that corresponds to either missing or zero values in Weather Normalized Source EUI field.

		2013			
		A	B	C	D
Office	Count	380 (83%)	45 (10%)	15 (3%)	17 (4%)
	Median	198.3	201.7	195.6	42.4
	Mean	204.4	230.4	320.1	2505.3
	Std	39.7	104.9	253.1	9295.4
	Min	79.0	92.6	74.4	4.3
	Max	342.8	411.3	957.4	38496.8
	Number of missing values	17 (4% of count)	12 (27%)	6 (40%)	4 (24%)

Table 4: Descriptive statistics of the results of DataIQ scoring based on source EUI for Office properties

As a robustness check, Table 5 presents number of observations within top and bottom range of EUI and Energy Star score for the entire data set by each grade. The percentage of these extreme values in total grade band sample is shown in figure 10.

Grade	2013			
	Top 5% EUI	Bottom 5% EUI	Top ESS	Bottom ESS
A	2	3	1	0
B	3	11	4	0
C	3	6	3	3
D	14	3	11	2

Table 5: Top / bottom values for EUI and Energy Star score by DataIQ grade for 2013 dataset

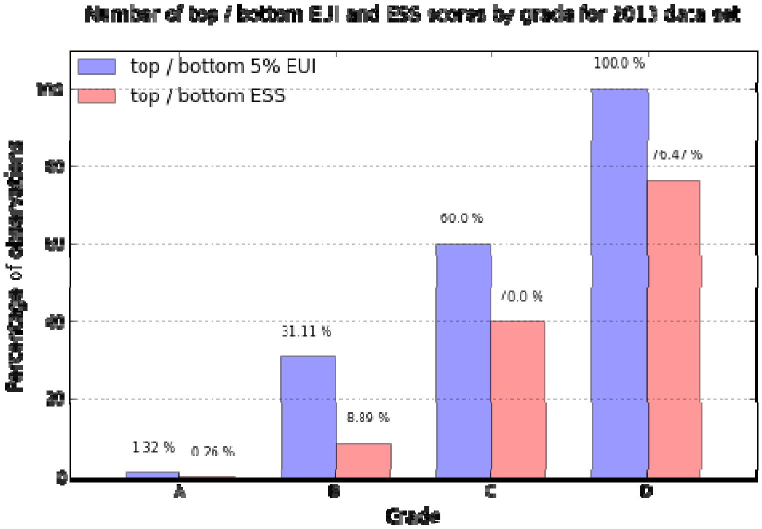


Figure 10: Top / bottom values of EUI and Energy Star scores as a percent of total observations per DataIQ grade.

The results shown in table 6 and figure 10 indicate the relative effectiveness of the DataIQ grade in identifying potential data quality issues. The above charts indicate that what might initially be viewed as outliers or unreliable data – those properties with EUI above the 95th percentile or below the 5th percentile – may not actually be a cause for concern. The DataIQ model provides additional guidance on which properties may be reporting questionable data or providing unexpected inputs, and thus provides a means for DDOE to better target outreach and auditing of energy disclosure data. The DataIQ algorithm provides a screening tool for understanding the relative quality and reliability of reported data that can be used to guide deeper, potentiality qualitative, examinations of such data. It also offers a measure of building energy performance that extends beyond simple metrics of energy intensity and the constrained approached used to calculate Energy Star scores.

It should be noted, however, that there are limitations to this approach. First, it has been designed as an easily-implementable model based on the rather limited scope of energy disclosure reporting requirements as it pertains to the full range of variables and characteristics that influence building energy use. Therefore, characteristics that have been found to influence energy use in buildings, such as construction type, systems information, occupant behavior, etc. are not included due to data limitations. We will explore opportunities to include such features in future iterations of the model to analyze the impact on prediction accuracy. Adding in data from additional cities could build a more widely applicable model. Second, the tool provides an in-sample, relative measure of reliability. Absolute measures of data accuracy were not available for the analysis presented in this paper. However, it may be possible to use data from buildings that have completed an energy audit as a training set to validate the model. Similarly, buildings that have third-party verified their Portoflio Manager inputs in order to receive the Energy Star label certification could comprise a more reliable training sample. On the other hand, these buildings are not necessarily representative of the building characteristics and energy use profiles of the broader range of buildings required to comply with energy disclosure ordinances.

DISCUSSION & CONCLUSIONS

Data quality has emerged as one of the primary challenges to extracting actionable insight from energy disclosure data. As with any self-reported data, errors in data entry, ranging from improperly entered data to incorrect data, can significantly undermine the validity of disclosure data analysis and its subsequent interpretation for policy. In addition, challenges emerge from an inability of some building owners to effectively and accurately collect key building characteristics that are necessary to measure relative energy performance.

Several areas of concern emerged with respect to data quality. First, questions and ambiguity regarding data definitions generated reporting errors in certain variables. Example of this include multiple buildings with a shared meter or buildings spanning more than one SSL. Second, manual input of energy data caused some discrepancies for those buildings that had not conducted an energy audit, received an Energy Star certification, or received whole-building aggregate data from the local utility. Third, for 2013 reporting, building owners were required to report on the space use and utility consumption data of *all* non-residential tenants including restaurants, gyms and other unscored tenants that had been previously exempt under unwritten Energy Star certification policy. The EPA Energy Star team has published this exemption and the DC Government has changed its guidance to also include this exemption for 2014 reporting. Finally, additional sources of data reliability concerns stem for default bias, data collection limitations (for example, many buildings do not track accurate worker density figures), and changes in reported data for the same building over time.

The analysis and methodology presented here provide an important step in improving the analysis of energy disclosure data and in identifying potential data quality issues with collected data. The DataIQ grade provides a composite measure of relative building energy performance and a leading indicator of potential data quality concerns. Although this is intended as a starting place, the DataIQ grade algorithm can be used to target potential outreach and expanded education and training for property owners and data providers and guide data audits and policy initiatives going forward. Future work will test the DataIQ methodology on energy disclosure data from other cities and explore a range of machine learning applications to examine the relative effectiveness of different models.

In addition, DOEE intends to use the methodology detailed in this paper to begin enforcing on data quality. The qualitative findings from this research on key drivers of data quality have already been incorporated into DOEE’s benchmarking guidance for the 2016. Moreover, DOEE intends to run the algorithms on all reports received in 2016 and beyond.

Buildings with low DataIQ scores will be flagged for follow-up. These properties may be subject to desk audits of the submitted data, with additional information and verification being required prior to being marked as “in compliance”. As shown in table 6, less than 10% of office building reports were given very low DataIQ scores, making the task of targeting these properties for enhanced review much more manageable.

In terms of broader implementation, DOEE and eleven other cities are using the the U.S. Department of Energy’s Standard Energy Efficiency Data (SEED) Platform™ to assist with managing the benchmarking data, matching it with other datasets, and sharing the results. The SEED Platform also provides a basic set of tools for identifying data gaps or erroneous values in reported data (Alschuler 2014). However, it is possible in the future that more sophisticated data quality analysis, of the sort discussed in this paper, could be integrated into or linked to the SEED Platform directly, greatly speeding the process of providing data quality review. As cities work iteratively with building owners and managers to improve the scope and quality of the data, the potential of benchmarking laws to drive energy savings will be greatly enhanced.

ACKNOWLEDGEMENTS

This work was supported by DC DOEE Grant # 2014-2-OPS, Project #5. We would like to acknowledge our sub-grantee, the Institute for Market Transformation (IMT), who conducted direct engagement with selected property owners as part of this grant. We would also like to thank the staff of DOEE and IMT for their helpful comments and feedback in developing this work. All errors remain the fault of the authors.

REFERENCES

Alschuler, E., J. Antonoff, R. Brown, M. Cheifetz. 2014. “Planting SEEDs: Implementation of a Common Platform for Building Performance Disclosure Program Data Management.” Proceedings of the 2014 ACEEE Summer Study on Energy Efficiency in Buildings in Asilomar, CA. Washington, DC: American Council for an Energy Efficient Economy.

Bellio, R. and Ventura, L., 2005. “An introduction to robust estimation with R functions.” *Proceedings of 1st International Work*, pp.1-57.

Capehart, B. L. and T. Middelkoop. 2011. *Handbook of Web Based Energy Information and Control Systems*. Lilburn, GA: The Fairmont Press, Inc.

Chandola V., A. Banerjee and V. Kumar. 2009. “Anomaly Detection: A Survey.” *ACM Computing Surveys (CSUR)* 41 (3): 15:1-15:58.

City of New York. 2012. *New York City Local Law 84 Benchmarking Report, August 2012*. New York, NY: Mayor’s Office of Long-Term Planning and Sustainability.

Hart, Z. 2015. “The Benefits of Benchmarking Building Performance.” *Institute for Market Transformation.* Washington, DC.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning: Data Mining, Inference and Prediction.* New York: Springer.

Hinge, A., D. Winston and B. Stigge. 2006. “Moving Toward Transparency and Disclosure in the Energy Performance of Green Buildings.” In *Proceedings of the ACEEE 2006 Summer Study on Energy Efficiency in Buildings*, 3:128-138.

Hsu, D. (2014). How much information disclosure of building energy performance is necessary?. *Energy Policy*, *64*, 263-272.

Keicher, C., J. Antonoff, B. Hooper, H. Beber, D. Pogue, and L. Cook. 2012. “Lessons learned from the implementation of rating and disclosure policies in U.S. cities,” *Proceedings of the ACEEE 2014 Summer Study on Energy Efficiency in Buildings*, 4:151-162.

Kontokosta, C. E. 2015. “A market-specific methodology for a commercial building energy performance index.” *The Journal of Real Estate Finance and Economics*, 51:288-316.

Kontokosta, C. E. (2013). Energy disclosure, market behavior, and the building data ecosystem. *Annals of the New York Academy of Sciences*, *1295*(1), 34-43.

Kontokosta, Constantine E. 2012. “Predicting Building Energy Efficiency Using New York City Benchmarking Data,” *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings*.

Palmer, K. L., & Walls, M. (2015). Can Benchmarking and Disclosure Laws Provide Incentives for Energy Efficiency Improvements in Buildings?. *Resources for the Future Discussion Paper*, (15-09).

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and buildings*, *40*(3), 394-398.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211-218.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, *12*(4), 5-33.