I²SL Benchmarking Working Group
Conference Call 5 February 2016

Attendees: Michelle Ruda, Alison Farmer, Dan Doyle, Tim Deak, Trina Innes (University of Alberta – Sustainability), Hadley Stolte, Rick Samson, Allison Paradise, Marcus Hazelwood, David Cohen, Joel McKellar

1) **Update on regression analysis** (Tim Deak)
   - Regression analysis helps us to understand what we can and can't do with lab data. We want to know whether we are asking the correct questions in our benchmarking efforts, and we want to understand the scatter in the data (which could be due to inconsistent interpretation of data needs by users, weather impacts, varying building efficiency, or other effects)
   - Tim re-ran the regression analysis done in ACEEE paper from ~5 years ago, using much larger current dataset (~500 cases).
   - As Tim described on a previous call, the adjusted R^2 value of the resulting correlations (11%) appears to be even smaller than before – there are real correlations, but there is a great deal of scatter.
   - Tim performed a set of checks on the dataset to verify that a linear regression analysis is appropriate (conclusion: model is valid).

   Testing Assumptions of Linear Model (notes from Tim):
   - *Residuals are not correlated—they are independent: Durbin-Watson around 2 (this is good)*
   - *Predictors are independent (not multicollinear): VIF is less than 10, not substantially greater than 1*
     - *Therefore, we can assume that the standard errors for coefficient values are accurate—but we see that the SEs for lab types are larger (and VIF is larger too)*
     - *The bs for labs are a bit less trustworthy because variance is larger—more variability across samples*
   - *Homoscedasticity of residuals: for each level of the predictor, the residuals of the outcome variable should have the same variance: a residuals plot shows some evidence of heteroscedasticity, but not severe.*
   - *Linearity: assumption met: the outcome variable should be linearly related to the predictors (we are using a regression line model, after all!)*
   - *Normality of errors: the errors are more or less normally distributed, but we have a large sample anyway—so even if this assumption were violated, it wouldn't matter much (due to the central limit theorem)*
   - *"Outliers" or "Extreme Cases" and their effect on the model. Do any cases have undue influence on the regression parameters?*
     - *Covariance ratio: all cases around 1, this is good*
     - *Mahalonobis Distances (distance of cases from means of predictors: with 552 cases and 6 predictors: no values are above 28 (also good)*
     - *Cook's Distance: no values are above 1 (also good)*
     - *Conclusion: no case unduly influences the regression parameters*
   - *Other bias in residuals:*
     - *The number of extreme cases falls within expected percentage of ±2 SDs for this sample size: 5%=23 cases (N=463) and we have 19.*
   - *Question (for Paul Mathew): I've dummy coded the lab type variable into several categorical variables. When running a regression with categorical variables, this compares each category against a control group (the uncoded category.) I'd like to see how this variable was entered into the original regression analysis.*
   - *Effect size: correlations between Site EUI and predictors (Pearson's r): all are below .3 – a small effect between each predictor and the outcome variable*

2) **Update on progress arranging access to Labs21 tool**
   - Best way to get access to tool and data is to have 3-4 people obtain LBNL affiliate status through I²SL. Two initial tasks will be assigned to affiliates.

- Task 1: Receive training on data validation and assist with entering backlog of new buildings into the peer group database. LBNL may receive funding for tool support this year, which would be a great opportunity for them ro provide us with training for future years.
- Task 2: Review Labs21 tool PHP code to determine ease of editing (e.g. is code commented?). This will help us understand the level of effort needed to make upgrades. Report back to group.
- Need volunteers to become affiliates:
    o Small amount of time commitment to become affiliate; could be more as the project moves forward. Make sure you can commit the time!
    o Data validation estimated to require ~40 hours/year total (not including training period).
    o If anyone knows PHP, please volunteer!
    o Volunteers so far: Alison Farmer; Hadley Stolte (non-expert PHP code review); Allison Paradise (data validation).
    o Request for volunteers will be sent out to rest of group too.

3) **The plan:**
- Can start to make changes to Labs21 tool once affiliate status is approved.
- We have a draft document summarizing intended scope of work. Once we have examined the tool's code, we can re-prioritize as needed.
- Proposed action items (in order of approach):
    o Website certificates problem – resolved (verified by Joel)
    o Enter 18-month backlog of new buildings into peer group database
        ▪ Count unknown (50-100 / year last few years)
    o Update documentation and instructions
        ▪ Can work on this now if we are looking for a task
    o Fix treatment of central plants
        ▪ Tool currently sums site energy regardless of type; this is an issue with chilled water usage in particular (and to a lesser extent with steam)
        ▪ Potentially make efficiency assumptions for central plants
        ▪ At least flag data
    o Add links to complementary I^{2}SL / DOE tools and resources
    o Update appearance
    o Make more historical data available
        ▪ Currently a single year is displayed for each building, supplanting previously submitted data. Need to check that LBNL still has the older data.
        ▪ Interest in this item – may warrant moving this up the list
        ▪ Alternative: Analyze historical data as a group and publish an article on the results, whether the data is added to the tool or not. Might be the simplest and most useful option.
    o Update graphing interface
        ▪ Make more intuitive
    o Community outreach to solicit more data and highlight new features (once we've made changes)
- Next Call – approximately 1 month from now.

Alison Farmer (afarmer@kw-engineering.com)
Michelle Ruda (mruda@brplusa.com)
2/7/2016