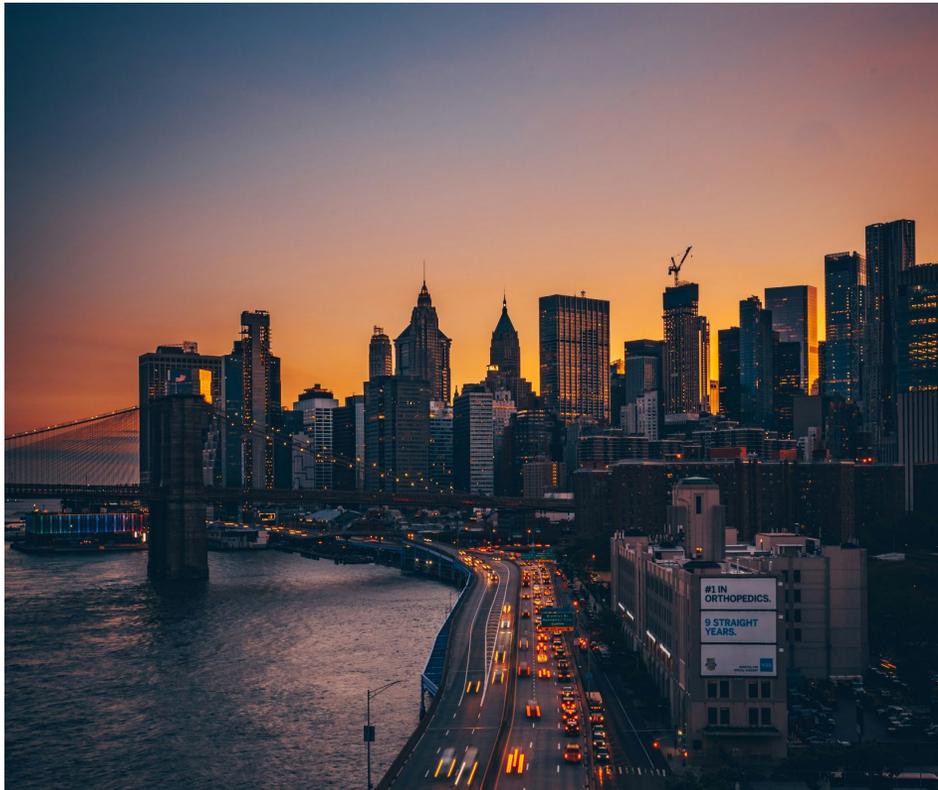# If the Shoe Fits

Statistical Techniques When Traditional M&V Methods Don't Cut It

Rahi Shah, Warren Energy

IEPEC

Energy Efficiency in New York City

Source: A city street at night by Lerone Pieters

- Highest density of high-rise buildings in western hemisphere

- $161M in Energy Incentives in 2024

- Accelerated Program

- M&V is triggered when:

    - High incentives or large savings

    - Projects proposing to install new technologies

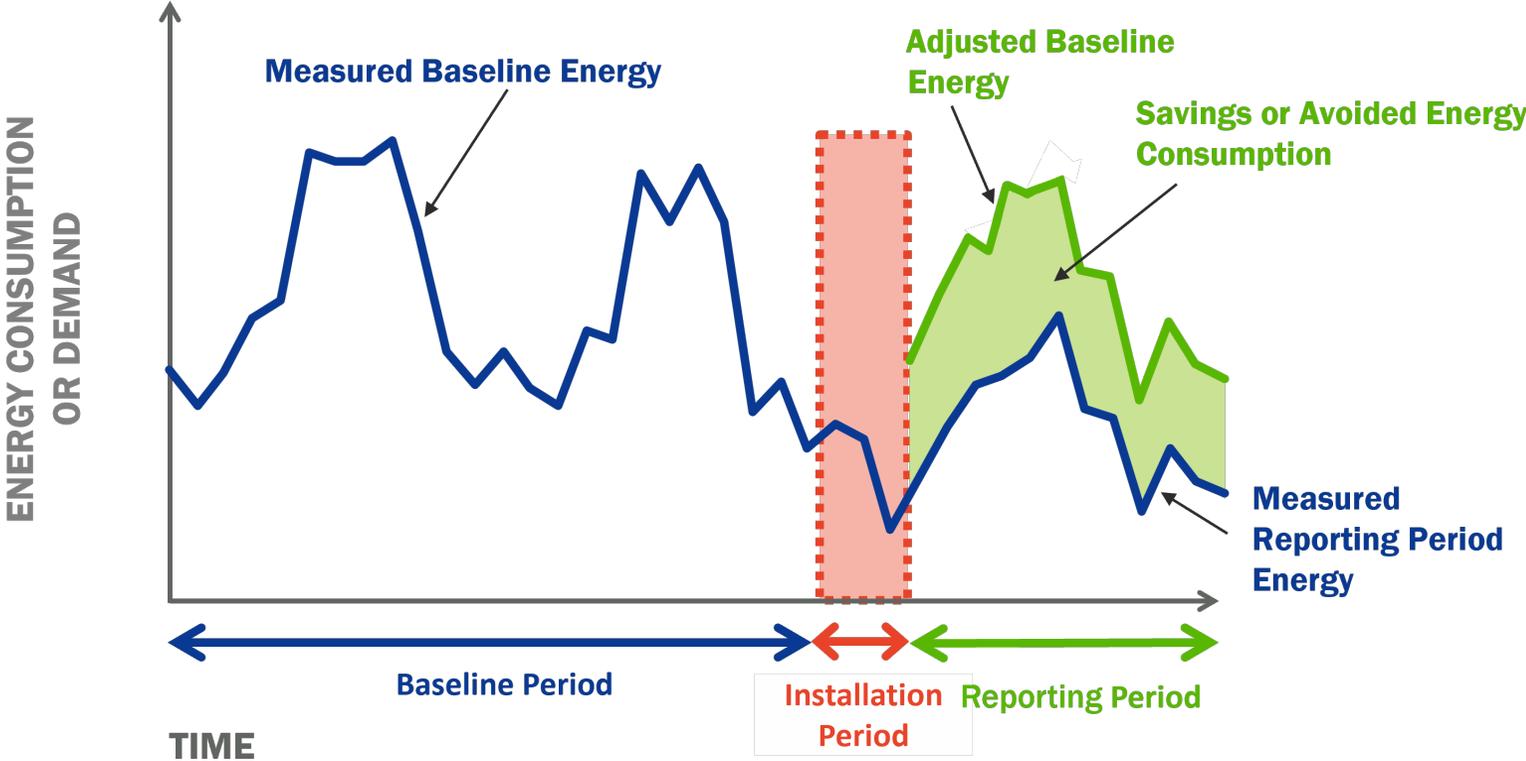   - Unique, complex, or risky applications
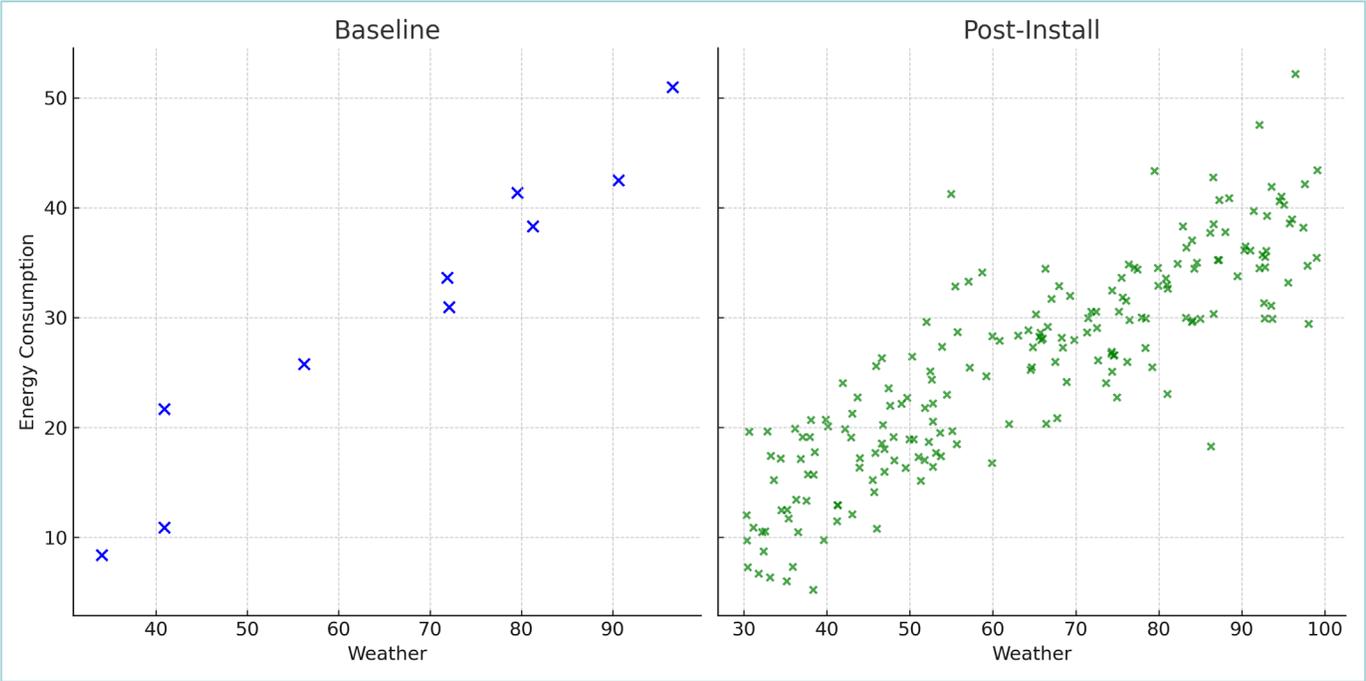
Traditional Techniques

And their limitations

Option C

ENERGY CONSUMPTION OR DEMAND

Measured Baseline Energy

Adjusted Baseline Energy

Savings or Avoided Energy Consumption

Measured Reporting Period Energy

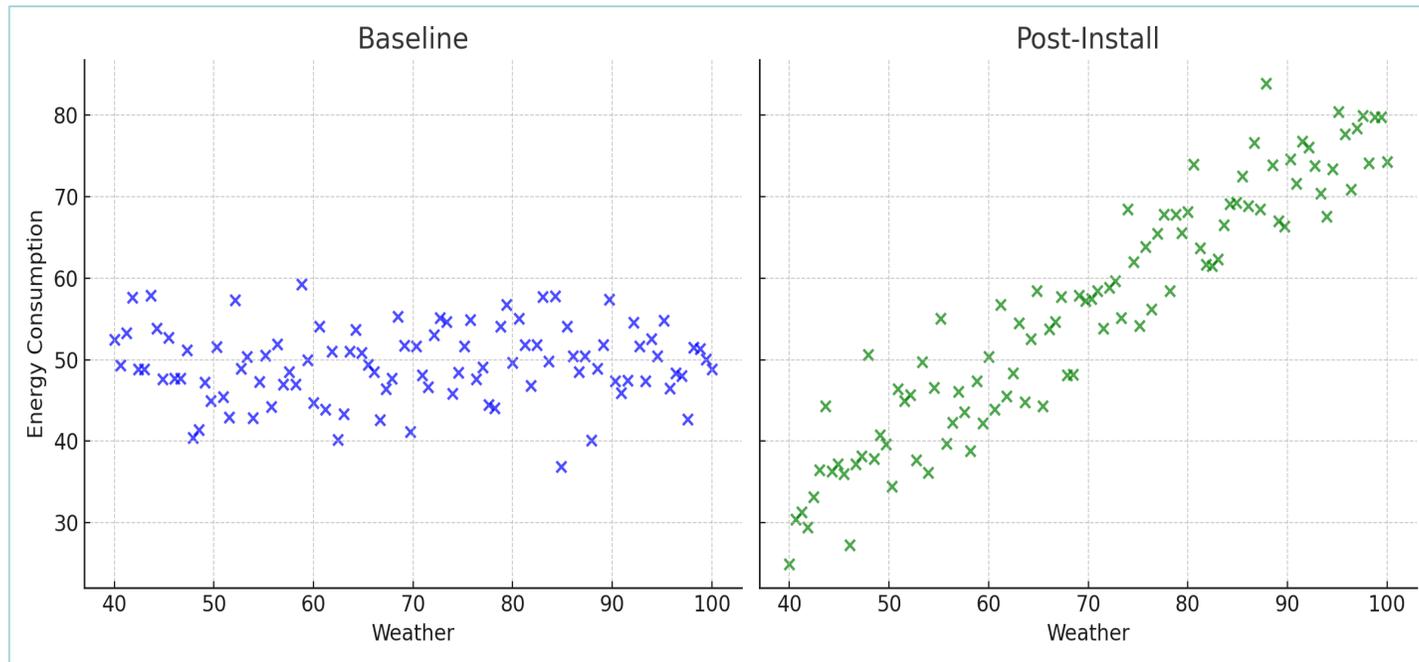Baseline Period

Installation Period

Reporting Period

TIME

Data Imbalance

Lack of strong correlation

- Multiple ECMs
- Non-routine events
- Savings must be significant (typically >20%)

Option A :
- New Technology
- Complex ECMs
- Interactive Effects
- Static and Simplified

Option D:
- Time and $$$
- Calibration can be tricky for existing buildings

Or...

Building Data

BUILDING OPERATION DATA

BAS / EMS DATA
- Supply/return/mixed air temps
- Fan speeds, damper positions
- Valve positions, reheat commads

ENERGY METERS
Whole building or subsystem energy data

WEATHER DATA
Outdoor air temperature, humidity

OCCUPANCY DATA
Sensors, scheduling systems

UTILITY BILLS
(OPTIONAL)

PROCESSING & MODELING
- Data cleaning & QA
- Feature engineering
- Baseline modeling
- Post-install data comparison

M&V ANALYSIS

ENERGY SAVINGS ESTIMATE

COST SAVINGS
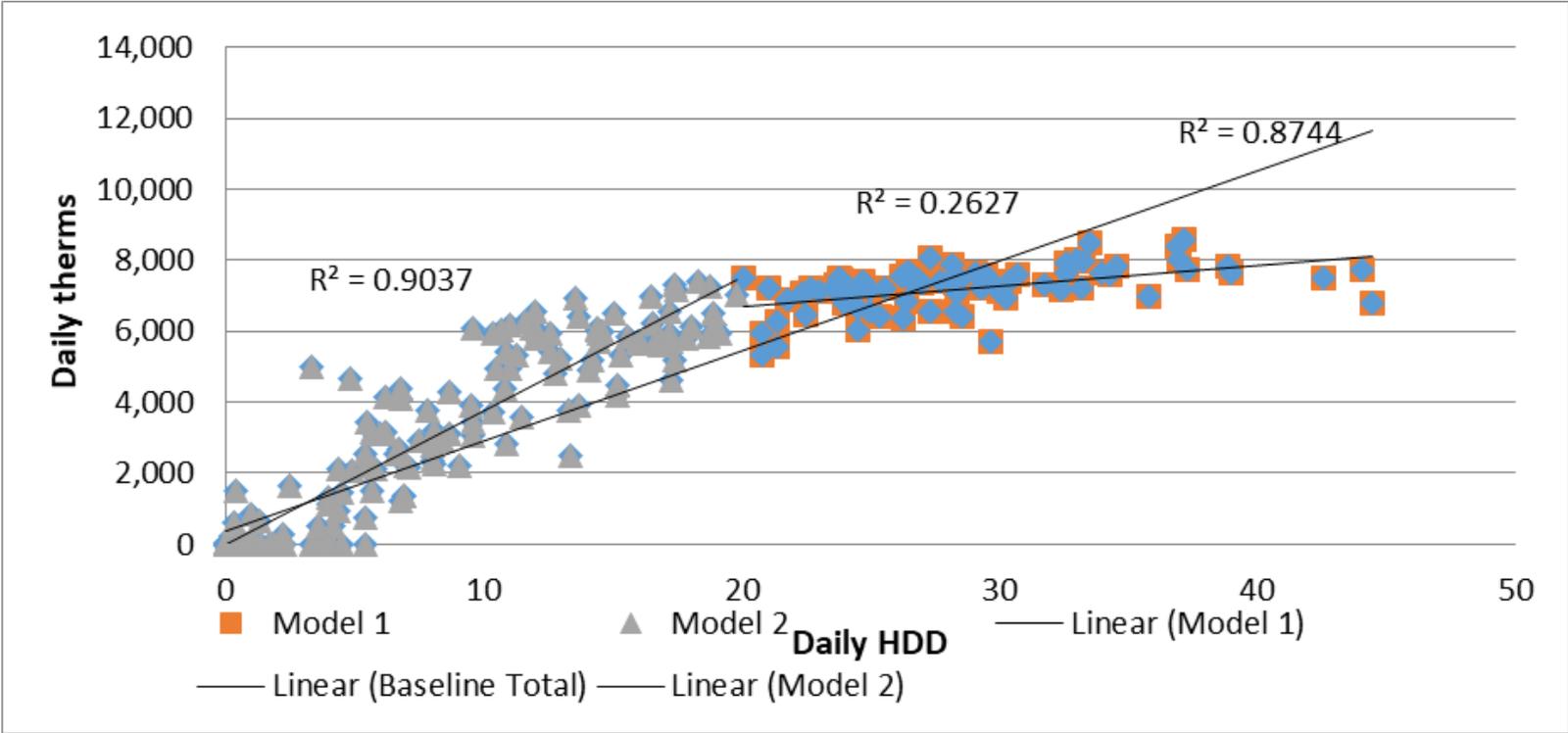
PERFORMANCE DASHBOARDS

REPORTS FOR UTILITIES / CLIENTS
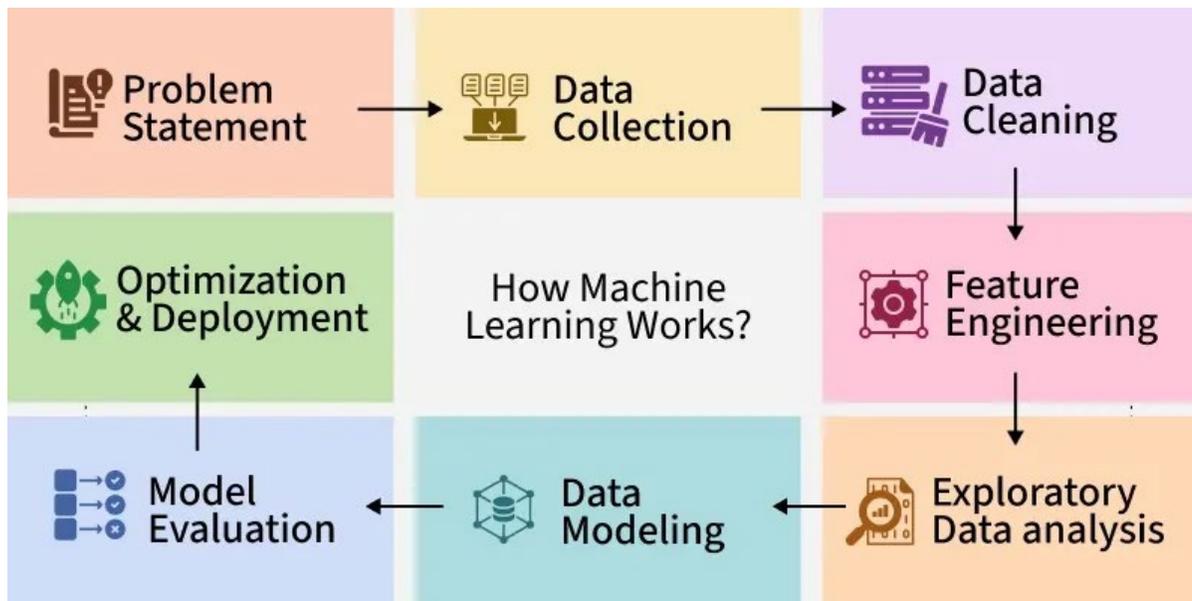
Statistical Techniques

Base Temperature Selection

Site	Pre-determined Base Temperature			Base Temperature Optimization			Change in Base Temperature (F)	Improvement (%)		
	R^2	CV (RMSE)	FSU	R^2	CV (RMSE)	FSU		R^2	CV (RMSE)	FSU
P	0.99	5.6%	28.3%	0.99	2.9%	14.7%	3	0%	48%	48%
Q	0.80	28.7%	26.9%	0.82	27.2%	19.1%	5	2%	5%	29%
R	0.95	7.3%	9.9%	0.95	5.9%	8.0%	3	0%	19%	19%
S	0.93	15.0%	53.4%	0.95	13.1%	42.6%	7	2%	13%	20%
T	0.95	25.9%	12.8%	0.95	25.7%	12.7%	2	0%	1%	1%
U	0.87	16.1%	44.0%	0.87	15.8%	43.7%	7	0%	2%	1%
V	0.91	14.6%	6.4%	0.93	13.8%	6.1%	7	2%	5%	5%
W	0.97	14.3%	29.2%	0.98	13.3%	27.3%	6	1%	7%	7%
X	0.96	13.0%	49.0%	0.98	8.8%	33.0%	5	2%	32%	33%
Y	0.95	18.8%	48.0%	0.97	15.2%	38.7%	6	2%	19%	19%

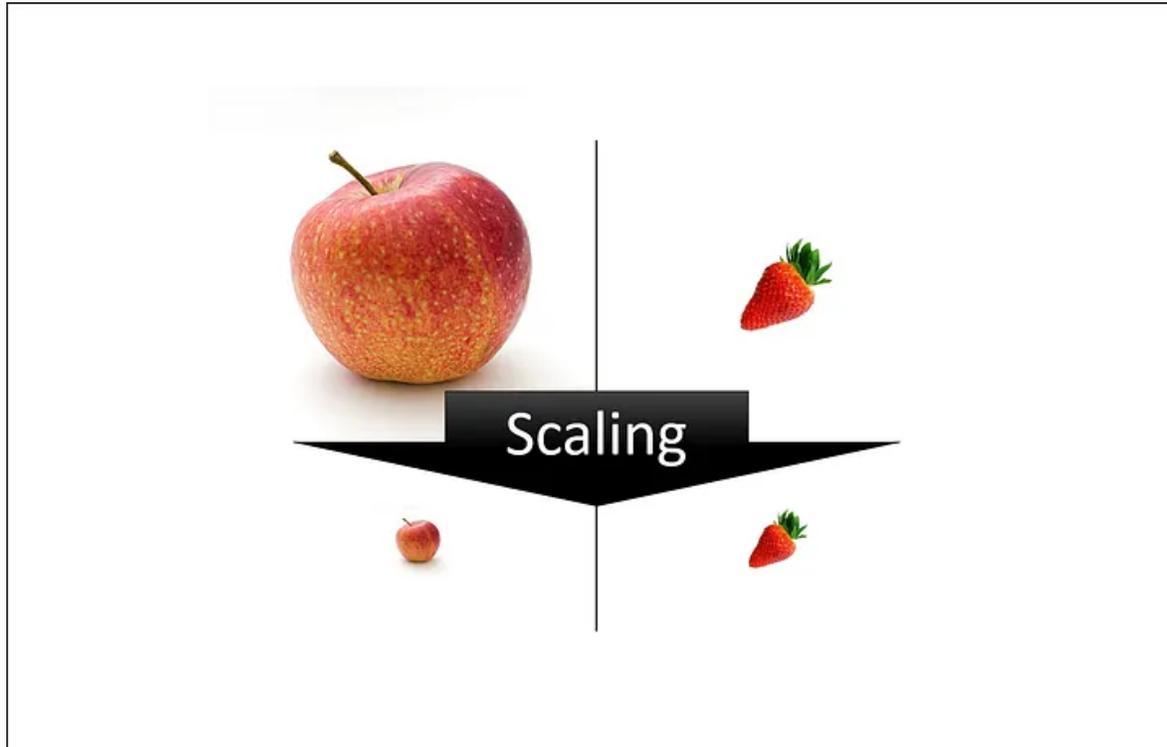| Con Edison Standard	
HDD	62
CDD	58.7

Linear Model Design
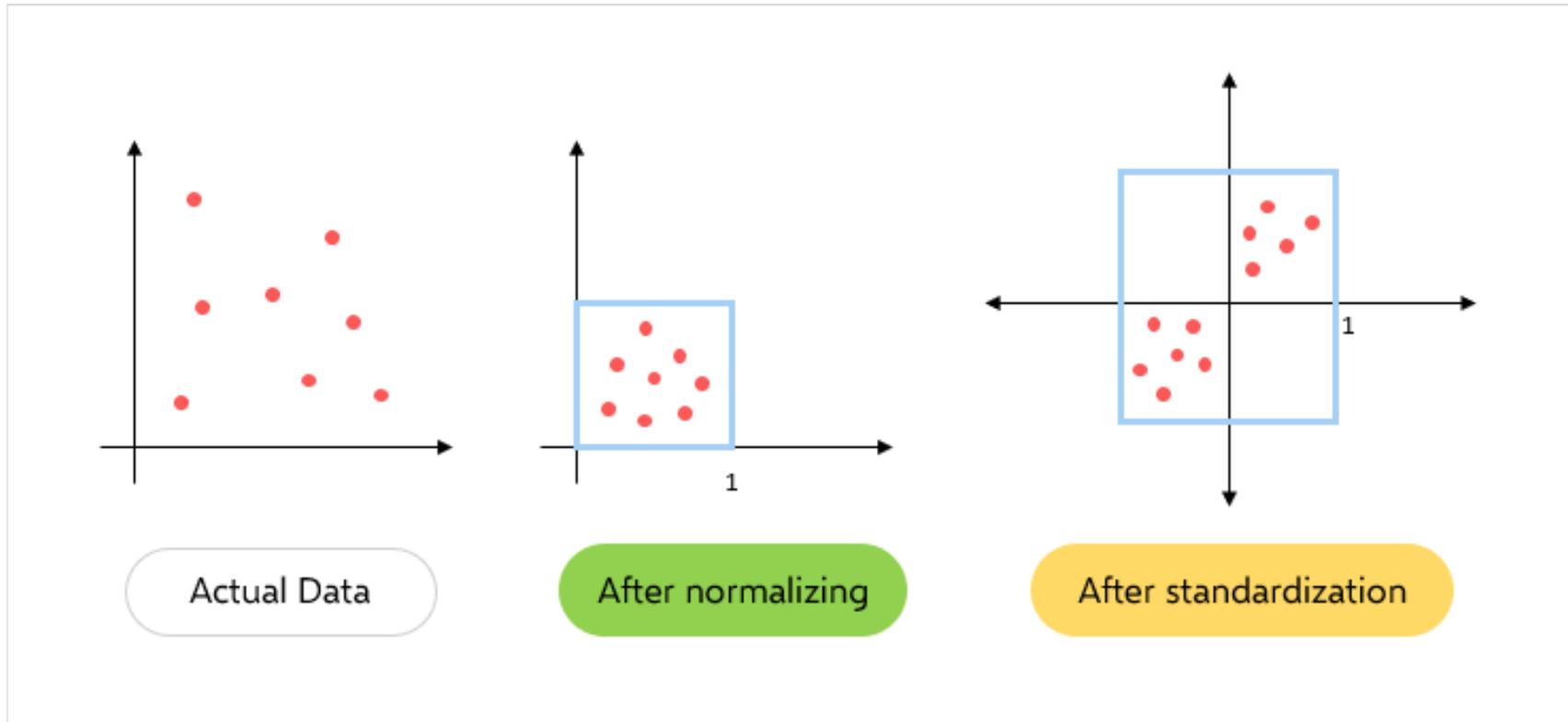
Machine Learning

Machine learning (ML) is a subfield of artificial intelligence that allows computer systems to **learn and improve from data without being explicitly programmed**

- Data Collection & Pre-processing

- Feature Engineering & Selection
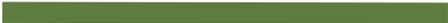
- Model Selection & Training

Feature Scaling

Feature Scaling

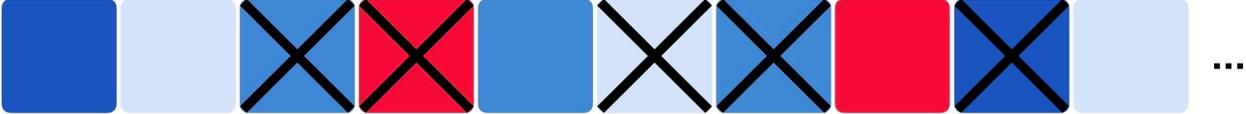Actual Data   After normalizing   After standardization

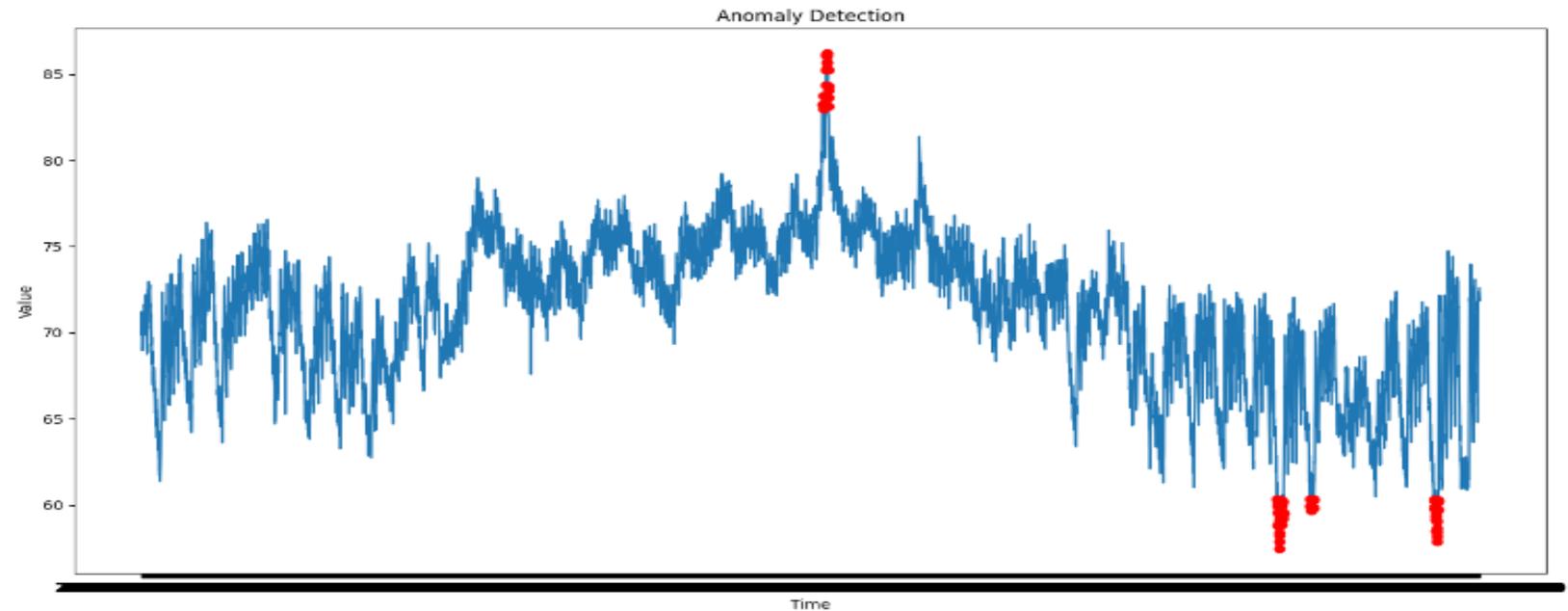# Feature Selection

All features

Identify informative features

Selected features

Anomaly & Non-Routine Events Detection

Anomaly & NRE Detection
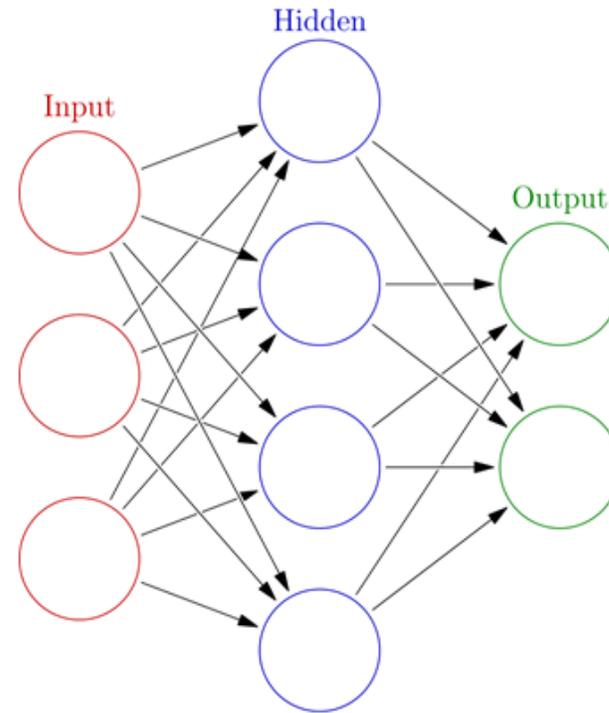## with multiple variables

Model Complexity

Case Studies

From Con Edison's C&I, MultiFamily and Clean Heat Programs

Site A: New Construction Clean Heat

Project Overview:

Location: Bronx, NY – multi-family high-rise development with shared amenities.

System Installed: Air-source VRF heat pumps replacing conventional fossil fuel systems → Eligible for **Clean Heat incentives.**

Baseline: ASHRAE 90.1 Appendix G code-minimum HVAC configuration.

.

Challenges:

Option C infeasible – new construction with no historical energy data.

AMI Data Limitations:

- Utility provided 15-min interval data.

- VRF compressor/condenser on shared meters; evaporator fans on tenant meters → difficult to separate heating vs cooling loads.

Changing **site occupancy** complicated weather-dependent regression models

Site A: New Construction Clean Heat

Data Processing Steps:

- Filtered out **unoccupied units** by setting weekly/monthly consumption thresholds.

- Detected and removed **outliers and high leverage points**.

- Integrated **occupancy rate** with weather variables as dependent inputs.

- **Scale normalization** applied to balance variables.

- Used **equipment efficiency curves + weather data** to project heating/cooling loads.

- Calibrated **Option A tool** at end-of-year occupancy rates.

Site A: New Construction Clean Heat

Site A: New Construction Clean Heat

Outcomes:

Accurate first-year savings **and future full-occupancy projections**.

28.4% correction vs. traditional engineering-based Option A estimates.

Corrected assumption that **cooling loads dominated**, influencing later tool revisions.

Demonstrated the value of **embedded submetering features** in overcoming conventional M&V limitations.

Site B: Hotel Smart Thermostat Retrofit

Project Overview:

Location: Midtown Manhattan, 518-room, 25-story hotel.

Existing System: Water Source Heat Pumps (WSHPs) without occupancy sensors → units ran **24/7**, wasting energy.

ECM: EMS-integrated **smart thermostats** with occupancy sensors linked to PMS for automated temperature setbacks.

- Recovery time kept **<20 minutes** when rooms became occupied.

Challenges:

Hotel had just reopened → **no stable baseline period** available.
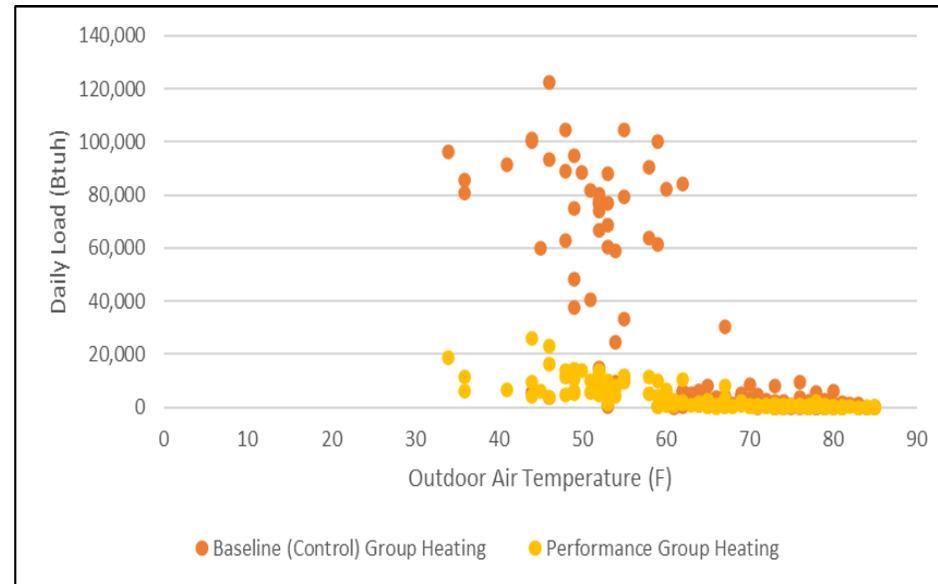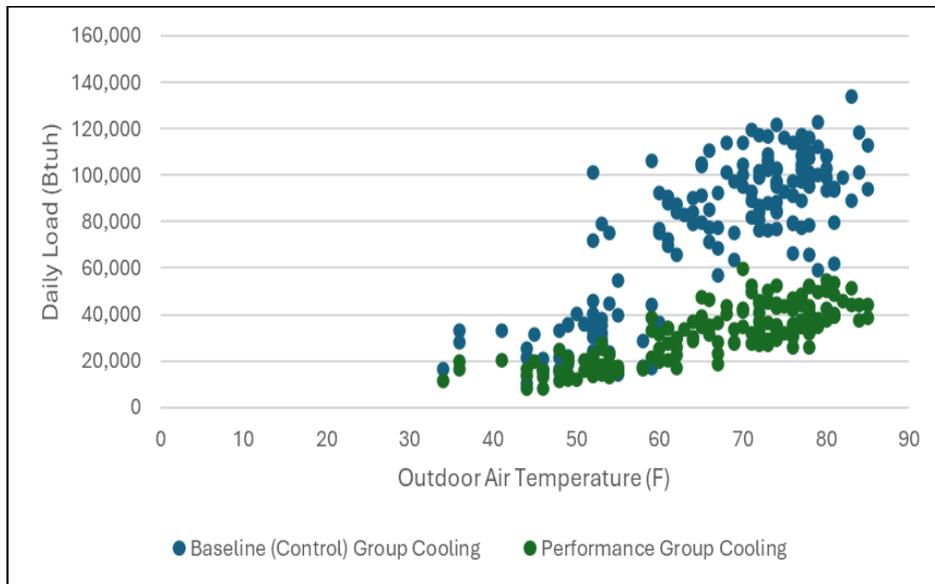
Occupancy climbing (85%+) → Option C infeasible.

Opportunity:

Smart thermostats had **trending & bypass modes**, enabling an **in-situ randomized M&V experiment:**

- 4 floors in **baseline/bypass mode** vs. 4 in **post-ECM mode.**

- Collected high-resolution data: room occupancy, fan/compressor run hours, setpoints.

Site B: Hotel Smart Thermostat Retrofit

Site B: Hotel Smart Thermostat Retrofit

Key Insights:

Expected savings during **unoccupied periods** were confirmed.

Unexpected differences during **occupied periods** revealed manual housekeeping adjustments affecting baseline rooms.

- Identified through thermostat data → **baseline adjustment applied**.

- Without this, savings would have been **overstated**.

Results:

Created **separate heating & cooling load profiles,** accounting for interactive effects across WSHPs.

Final verified savings:

- **102% higher** than applicant estimate.

- **~350% higher** than conservative engineering estimates

Site C: Manhattan High-Rise
Fluid Treatment System Evaluation

Project Overview:

Site: 27-story commercial high-rise in Manhattan, NY.

Goal: Evaluate new **fluid treatment system** projected to improve **chiller efficiency by ~15%.**

Challenges:

No functional BAS or baseline system performance data.

Chiller plant **not submetered** and **variable occupancy patterns.**

Low savings potential (<10%) → Option C not viable.

Site C: Manhattan High-Rise
Fluid Treatment System Evaluation

Solution:

Operated system in **bypass mode for 3 months** to collect baseline data.

- **Chiller amps, BTU meters**, entering/leaving water temps.

Developed **custom baseline energy model** to capture performance more accurately than standard bi-quadratic curves.

Modeling Methods:

Compared **Gordon-Ng thermodynamic model** vs. **data-driven models**.

Random forest regressor provided highest accuracy ($R^2 = 0.92$).

Used **feature selection** to identify key drivers:

- Chiller plant tons

- Chilled water return temperature

- Condenser water supply temperature

Site C: Manhattan High-Rise
Fluid Treatment System Evaluation

Outcomes:

Empirical models **outperformed thermodynamic approach by 22%** (cross-validated R²).

Baseline model predicted post-retrofit consumption → **avoided energy savings calculated and weather normalized.**

Demonstrated that **advanced modeling + feature engineering** can accurately capture performance for projects with **limited data and small savings signals.**

Model Performance Parameter	Gordon Ng	Linear Regression	Decision Tree
Model R-squared value (cross validated)	0.72	0.85	0.88
Model R-squared value (test set)	0.70	0.82	0.85
Model RMSE-CV	19.5	14.76	13.12
CV-RMSE predicted vs. actual	0.24	0.05	0.03
R-square predicted vs. actual	0.63	0.83	0.92

Key takeaways

Embedded Data Boosts Accuracy: Exploit all data resources available on hand

Creative Baselines: Bypass periods, synthetic models, and randomized experiments can overcome lack of historical data.

Data Cleaning & Feature Selection: Removing outliers, filtering unoccupied units, and selecting key variables can vastly improve models.

Advanced Modeling (Sometimes) Matters: Advanced models & regressions can sometimes capture complex interactions better than traditional methods. Caveat: Loss of interpretability

Combining embedded data, smart baselines, and advanced models enables **high-confidence M&V** for complex buildings.

Thank you

Rahi Shah

rahi@warren-energy.com

www.warren-energy.com

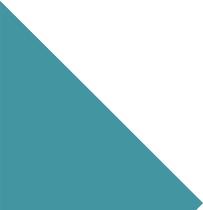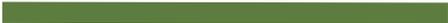# Appendix

Feature Selection in Energy Models

Why it matters:

- Too many variables → **overfitting** and **reduced interpretability**.

- Proper feature selection improves **accuracy** and **insight** into savings.

Feature Selection Methods:

- **Filter Methods:** Drop features with low correlation (e.g., ANOVA F-test, Chi-square).

- **Wrapper Methods:** Evaluate subsets and remove less impactful features (e.g., recursive feature elimination).

- **Embedded Methods:** Feature selection during model training (e.g., Lasso, Ridge, Gradient Boosted Trees).

- **Dimension Reduction:** PCA for simplifying complex, high-dimensional data.

Anomaly & Non-Routine Event (NRE) Detection

Why it's critical:

- NREs **inflate or deflate** savings estimates, impacting confidence intervals and FSU values.

- Examples: Missing utility readings, sensor errors, leaks.

Detection Techniques:

- **Visual Review:** Time-series plots, site inspections, interviews.

- Statistical Tools:

  - **Studentized Residuals:** Outliers usually outside -3 to +3.

  - **Leverage Statistic:** Identifies data points far from mean.

  - Plotting both together is a powerful diagnostic.

- **Advanced Methods:** PCA, k-Nearest Neighbor, SVM for high-dimensional datasets.

Scaling, Time-Series & Advanced Modeling

Normalization:

- Needed when features have **different units/ranges** (e.g., degree days, occupancy %).

- **Z-score normalization:**
$$x' = \frac{x - \mu}{\sigma}$$

- Ensures fair comparison and stable regression.

Scaling, Time-Series & Advanced Modeling

Key Time-Series Practices:

- **Cross-Validation:** Prevent overfitting by testing on subsets of baseline/post data.

- **Autocorrelation Checks:** Use tests like Durbin-Watson; mitigate with lag terms or differencing.

- **Anomaly Detection:**

  - Z-score or seasonal decomposition (univariate).

  - Hotelling's Distance, PCA, SVM (multivariate/non-linear).

Scaling, Time-Series & Advanced Modeling

Modeling Trade-Offs:

- **Linear models:** Simple, interpretable.

- **Non-linear models:** Higher accuracy but risk of overfitting and lower interpretability.