Misapprehension about Configuration and Artificial Intelligence Adoption Rates in Building Automation

Paul Campbell
Vice President of AI & ML Solutions, facil.ai

Contents

Objective

Endemic misapprehension about the very nature of data, parameters, and configuration and how misunderstanding of their interactions within the Building Automation space - specifically how they can inform performance and value gained - likely influences the slow Artificial Intelligence adoption rates observed over time given the technology milestones already achieved. Relieving this misapprehension is critical to achieving the fullest value from Building Automation assets. Artificial Intelligence provides the means to do so but only if the foundational understanding of configuration has already been attained. In light of the advance of Artificial Intelligence and the technology horizon before us, foundational understanding of configuration holds larger implications than ever before.

The Hierarchy of Data, Parameters, Ruleset and Configuration and its Importance

Like any hierarchy, at the base are all the data in need of organization (classification, tags, etc.) and control, which are provided by parameters and the governing ruleset as the data's transit ascends the ranking elements in the hierarchy. Together these elements comprise an immutable object specifically designed to control in a manner detailed in the ruleset.

Volume and Amplitude in the hierarchy have an inverse relationship: where we find greater volume we will see lesser amplitude and vice versa. Volume for our purposes can be understood to be a vast ongoing collection of the most granular observed or measured data in the form of time-value pairs. Amplitude can be understood to be where the strategic directional control is exercised, where a change would directly alter the core behavior and performance of the system(s) concerned. In between where volume and amplitude are in close proximity is where amplitude is expressed, most commonly through functions that are guided by parameters.

<table>
<tr><th colspan="6">Immutable Architectural Object</th></tr>
<tr><td rowspan="8">Configuration</td><th>Element & Hierarchical Position</th><th>Examples</th><th>Volume</th><th>Amplitude</th><th>Rates of Change</th></tr>
<tr><td rowspan="2">Ruleset (Top)</td><td>Authentication</td><td>Low</td><td>High</td><td>Low</td></tr>
<tr><td colspan="4">Definition: A defined, organized, and evaluable collection of policies, conditions, and actions that dictate the specific computational logic and behavior of a system or subsystem in response to incoming data.
Area of Control: The evaluation of data and the application of parameters.</td></tr>
<tr><td rowspan="2">Parameters (Middle)

μ (mean), σ (standard dev)</td><td>Set Points</td><td>Moderate</td><td>Moderate</td><td>Moderate</td></tr>
<tr><td colspan="4">Definition: A specific, named variable acting as a structural element.
Area of Control: The operational logic on the data.</td></tr>
<tr><td rowspan="2">Data (Base)

χ (for a single observation)</td><td>Time-value pairs</td><td>High</td><td>Low</td><td>High</td></tr>
<tr><td colspan="4">Definition: The input, subject to variability with each observation.
Area of Control: The subject of the operation, independent of hierarchy.</td></tr>
<tr><td colspan="5">Definition: The collection of all settings and their assigned values.
Area of Control: The overall environment and rules.</td></tr>
</table>

Table 1: Configuration as a Designed Object

Many have an awareness of the hierarchy but perhaps not as many have a proper understanding of it. A parameter is not limited to a fixed value but holds the possibility of a conditional range of values. A conditional range created by a ruleset renders the flexibility and the agility that results in more refined control that extends value beyond that provided by the "80/20 rule" which we routinely accept and rarely question. The "80/20 rule" suppresses the maximum value gained capacity of a fixed value parameter by leaves substantial headroom where there is untapped value. It's a good place to start but not a good place to reside.

The Need for Agile Configurations

With a proper understanding of the hierarchy of configuration comes the knowledge that flexibility is the capacity to change, and agility is the speed and efficiency with which that change occurs. Part of that understanding is knowing that there are the implications when changing elements of a configuration, summarized in **Table 2**.

Hierarchical Level	Change Type	Implication for Configuration
Parameter	The value of a parameter	Configuration is stable: the change is expected and managed by system design.
Configuration	The definition or range of a parameter	Configuration is changed: a fundamental rule or boundary of the system has been altered.
Aspects Concerned	**Flexibility / Agility**	**Implications**
Code Base	Not a design feature	Extending or altering core functionality and operational range requires modifying the source code.
Parameters & Configuration	A design feature	Reasonable core functionality and operational range was made extensible by design.

Table 2: Implications of Agility and Changes in Parameters

The present workplace places a premium on operational and cost efficiencies. Ultimately this compels one to seek solutions for optimization of the building automation assets but often this requires a mindset shift when it comes to configuration. A realistic approach is to conceive a main line of defense (the dash-dot arc line scribed across **Figure 1** and seek a suite of configuration solutions for units on either side of the main line of defense. It is no coincidence that this aligns with the consensus of industry professionals in **Table 3**.

The Cluster Analysis in **Figure 1**Error! Reference source not found. illustrates a common happenstance regarding variation in performance, even within narrowly defined population groupings such as geography, age, prototype, etc. Many experience this but few have the support to address it, at least by conventional means of support. There will always be some variation of performance but only through a thorough understanding of the hierarchy and a commitment to overcome prevailing technology constraints will it be possible for an intervention regime to keep units demonstrating high probability from descending into the next lowest cluster in a systematically automated way thrive.

A word about variability in performance: it is perfectly natural for variability to occur. Even after the most advanced solution is deployed data will inevitably undergo a re-clustering, and the current frontier of outliers will be revealed. This is to be expected. It is not the clusters themselves but the high probability of risk we seek to solve. Technology solutions should therefore be thought of in terms of evolutionary change.

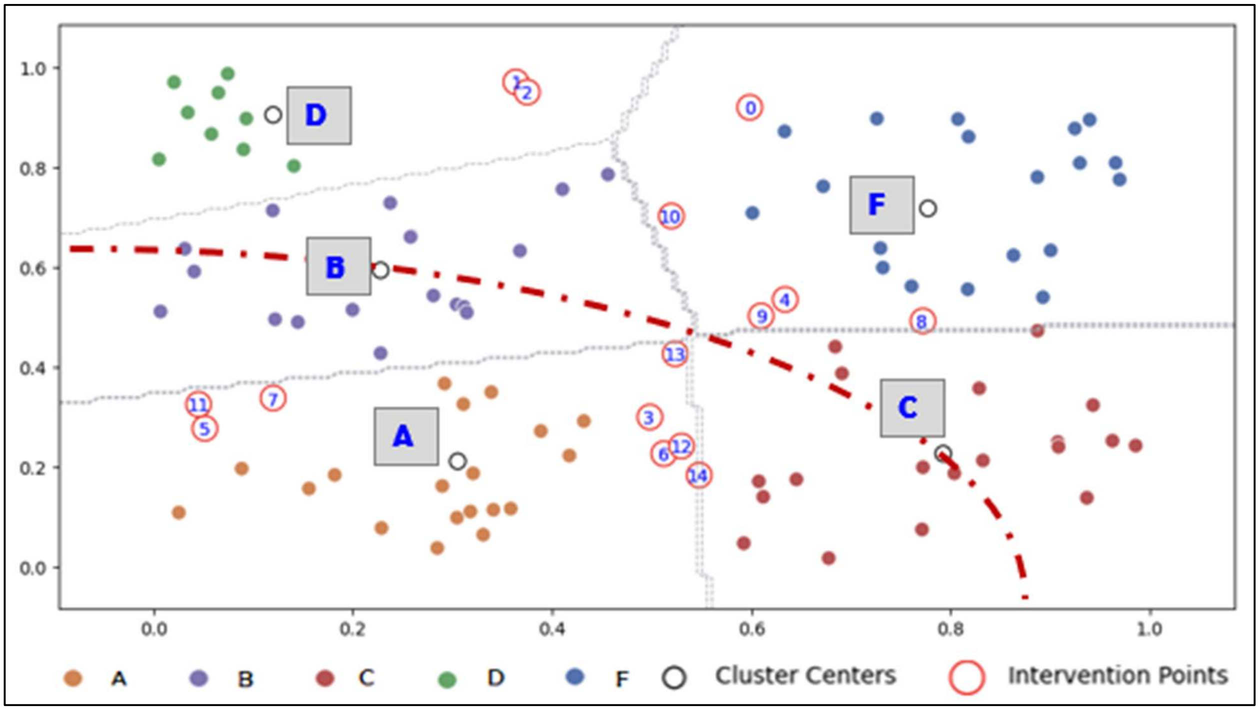

Figure 1: Generic Cluster Analysis of Randomized Values demonstrating range of performance

Table 3 gives us a sense of scale by displaying a demographic survey summary of the estimated number of units in the field. Per the EIA, there are approximately 1.1 million buildings in the US that are not in the replacement category or in the end of life stage. Consensus among industry professionals articulates their need for more robust optimization in their BAS fleets. Assuming an even distribution of these 1.1 million facilities in our randomized Cluster Analysis, we can see the need for any solution under consideration to persist in a highly variable environment. Artificial Intelligence is the obvious way forward.

Release Year: 2021-2022 | Data Year: 2018						
Age Band	**Replacement**			**Non-Replacement**		**Notes[1]**
	Rates	**Min**	**Max**	**Min**	**Max**	
>= 16	30%	252,300	258,300	-	-	End of Life risk
8 - 15	45%	378,450	387,450	**462,550**	**473,550**	Needs functional modernization
0 - 7	25%	210,250	215,250	**630,750**	**645,750**	Needs optimization
Total	100%	841,000	861,000	**1,093,300**	**1,119,300**	

Table 3: Summary of US Energy Information Administration Commercial Buildings Energy Consumption Survey.
[1]Notes reflect a consensus of the Building Automation System and Facilities Management (FM) industries.
Source: Detailed Table B1:

Paul Campbell
Vice President of AI & ML Solutions, facil.ai

Role of Artificial Intelligence in Configuration

Reading **Figure 2** and **Figure 3** in aggregate, the mainstream building automation industry technology largely exists at a 1990 level state of the art, which translates to the Lagging Vanguard based on the timescale of AI accessibility. That implies AI has headroom to accomplish more.

Observations on Evolution of AI

Tracking the evolution of technology (**Figure 2**) as well as it adoption rates (**Figure 3)** and reading the **Table 3** notes reveals an interesting context about the present state of those legacy units in the field with long lifespans remaining: a happy medium seems to exist well after the time when tried and true technology was propagated on an 80/20 rule basis and was procured when this compromise was absolutely necessary to achieve the then possible level of automation. Technology has evolved and this acceptance is being challenged by present-day needs.

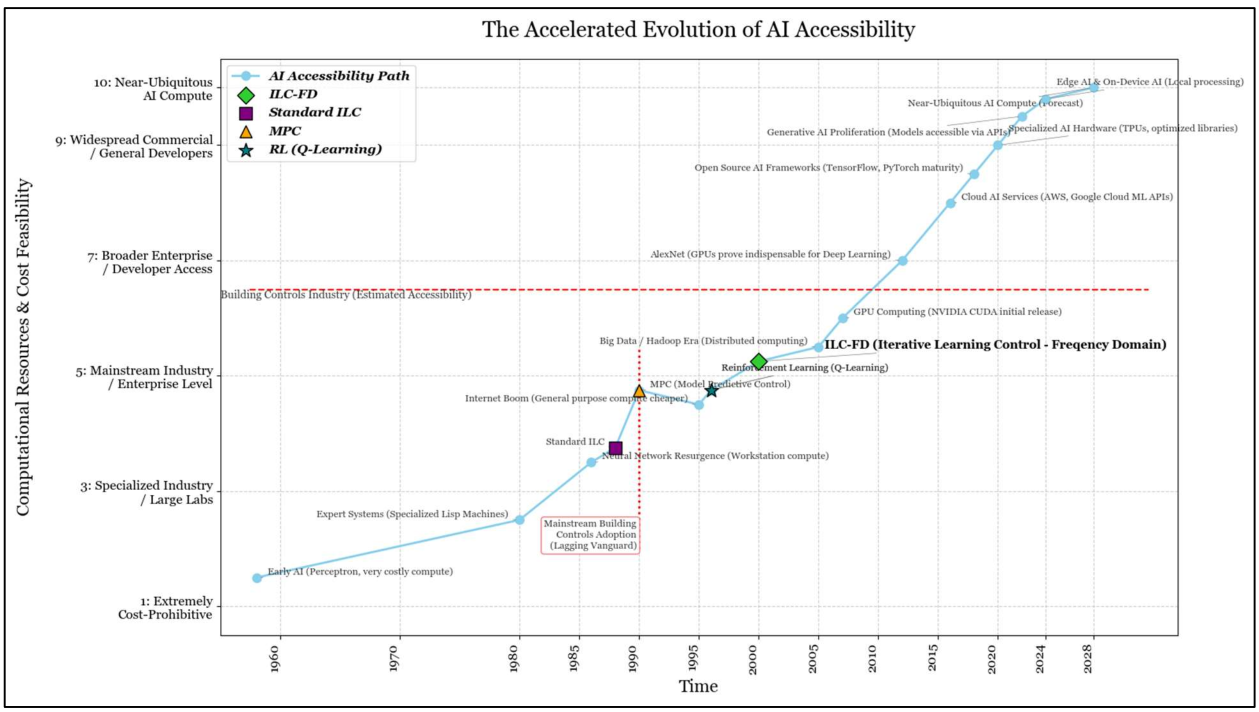

Figure 2: Evolution of Artificial Intelligence over Time

Observations on AI Adoption Curves over Technology Milestones

However, given the demography and rates of adoption it is not surprising nor does it reflect poor business practice that the Lagging Vanguard reflects circa 1990 innovative technology. The Lagging Vanguard is fine if one accepts operating only within the 80/20 rule. Industry consensus from **Table 3** suggests more is required to penetrate the unoptimized 20% than out-of-the-box technology permits or that well intentioned professionals with imperfect understanding of the nature of a configuration hierarchy can effectively manage if left to their pathognomonic impulses can secure.

Generally speaking, increased optimization exists just beyond the Lagging Vanguard horizon within the Early Majority state of the art, and this state reflects a leap forward in technology milestones that plots in the 2005-2012 timeframe, a leap of 15 to 22 years. Getting to this Early Majority state of the art is paramount to meet present day requirements.

It is unrealistic to expect a vast population to just leap straight to the most innovative technology at the innovator stage of development and bypass the testing and piloting that plays out over a number of years. That is why adoption curves exist in any large endeavor and help define the durable nature of commercial grade products. What is being suggested here is that the milestones of Artificial Intelligence that lay immediately ahead on the conceptual time phase per **Figure 3** tend to straddle the Early Adopters and Early Majority curves and that the time to recalibrate our expectations for optimization is already at hand.

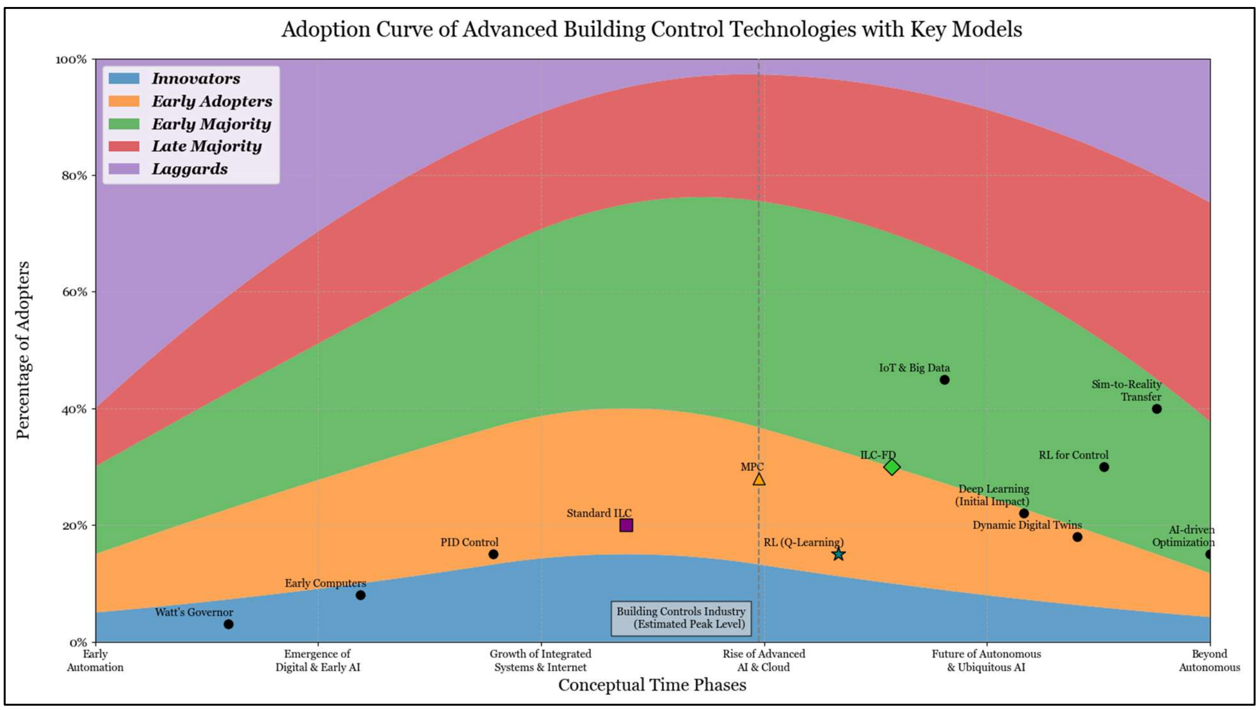

Figure 3: Artificial Intelligence Adoption Rates over Time expressed as Technology Milestones

This brings us back to configuration and the industry consensus: there is a need not currently being met given technology limits on configuration. More advanced forms of Artificial Intelligence suggest it is possible to augment the reach and performance of configuration when viewed through the lens of optimization.

The psychrometric graphs in **Figure 4** and **Figure 6** illustrate how optimized configurations via Artificial Intelligence solutions for RTUs and Chillers compare to configurations that are not optimized. The intent is to use linear distance and boundaries as tools to illustrate operational area and lend context to how a persistent optimization paradigm performs in an environment subject to variability, which can be expected widen over a large enterprise.

AI in Action: The Psychrometric Views of RTU Efficiency Configurations

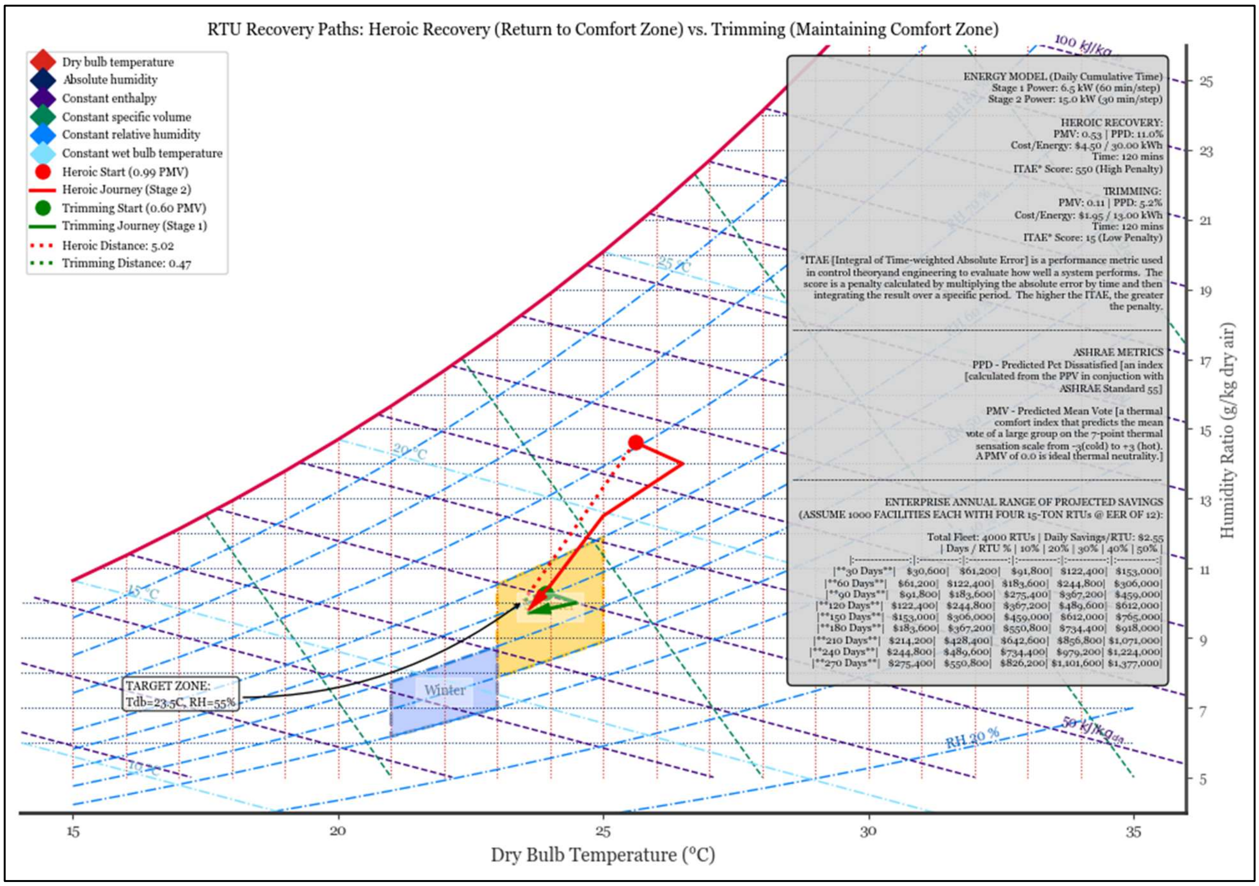

Figure 4: RTU Psychrometric Chart

Observations

This psychrometric chart for an RTU displays two configuration journeys, one for heroic recovery requiring two stages of cooling where configuration allows for the temperature to drift out of the comfort zone during unoccupied hours; and one for trimming, where configuration prevents excursion from the comfort zone relying heavily on one stage of cooling on a basis approaching exclusivity, using the second stage minimally and only as a peaking asset.

Euclidean Distance and the Fallacy of Unit Run Time as a Measure

The Euclidean distances for **Heroic Recovery** and **Trimming** (dotted lines on the psychrometric chart) are calculated to visualize the impacts of the two configurations and to contextualize the differences between them. Total run time is too broad a measure of performance and is only useful if parsed into distinct run times per stage. Common understanding on run time as a measure has it that "more is bad and less is good" and that run time for both stages are of equal value. This is a misunderstanding, and it leads to poor configuration with negative impacts. The impact compounds if this misunderstanding propagates over an enterprise's fleet of RTUs. If one understands that each stage of cooling represents an additive component of a building's energy demand and that the main line of defense is to limit the use of the second stage of cooling as much as possible - even at the expense of increased run time of the first stage of cooling – the net will be improved energy efficiency and less energy used.

There is an energy industry parallel at the grid level that drives this point home but is understood perfectly well: a more expensive energy generation peaking asset is brought online only as necessary as baseline assets approach their capacity to meet system-wide demand. In this context, there is no trouble understanding that a peaking asset is not co-equal to a baseline asset and so it should be with RTU cooling stages.

Control Performance Measurement: ITAE and Value Gained

Error: Valuation:

$$ITAE = \int_0^T t * |e(t)|dt \qquad Log\ Value\ Gained_k = \ ln\left(1 - \frac{ITAE_k}{ITAE_{Initial}}\right) + \ \epsilon$$

Equation 1: Error and Valuation in the Time Domain

The ITAE (Integral of Time-weighted Absolute Error) is a metric used extensively in control theory to evaluate the performance of a control system's transient response, gauging how quickly and smoothly it reaches a new setpoint after a disturbance. By multiplying the absolute error by time, the ITAE equation ensures that errors occurring later in the response are penalized more heavily than errors occurring immediately as illustrated upon the convergence of Zone Temperature data in the persistence phase in Figure 9. This compels a solution that prioritizes speed. The Value Gained is simply the log of the inverse of ITAE error.

Using set point as the lever it was always intended to be plays into a good ITAE score. One must understand that a set point is a parameter and not data. It holds no intrinsic value. Although it may be expressed as a number with a unit of measure, it is just a numerical representation. However, the set point does hold operational value as a lever within control theory that impacts the ITAE score.

For example, configurations based on a fixed 73*F occupied cooling set point and a 78*F unoccupied cooling set point, may not fit the needs of a building during the warmer months of the year as there will be more heroic recovery traversing longer Euclidean distances back to the comfort zone resulting poor ITAE scores relative to an optimized configuration that limits stage 2 cooling to its minimum. Control in the former configuration is often inadequate during warmer months with increased error and diminished value gained.

If we trended the data on a dynamic configuration based on Artificial Intelligence, we would see a convergence where control is expressed and is persistently tuned, at which point ITAE Error is weighted more heavily and Value Gained increases logarithmically.

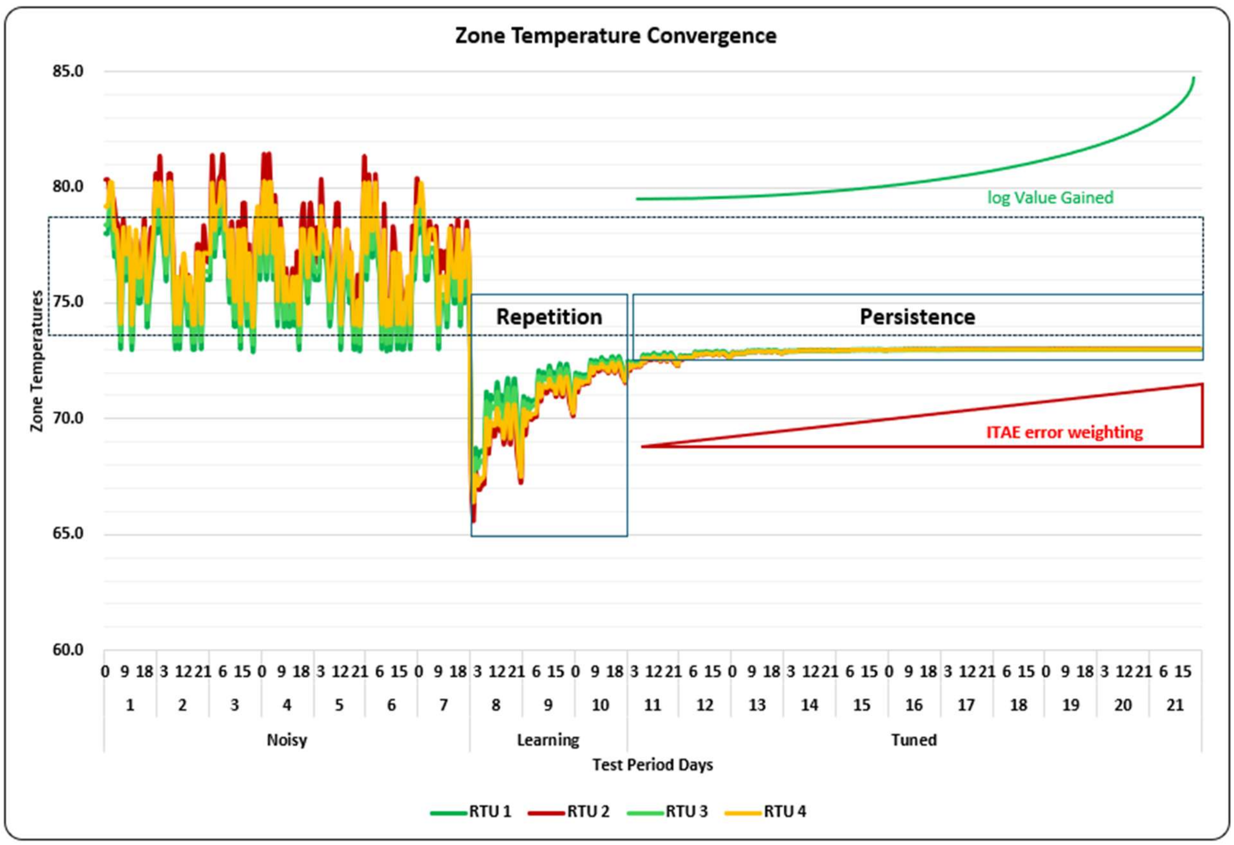

Figure 5: Convergence with Error Weighting & Value Gained

Not every configuration hits the same, in terms of trying to quantify the Error and Value Gained metrics of the underlying model, which may operate in substantially different manners. It should be understood that these model performance metrics are distinct from common business metrics like ROI and Payback Period, etc. Error and Value Gained are metrics that measure control efficacy, but domain must be considered. Model Predictive Control for instance should have Error and Value Gained metrics rendered in the Time Domain while ILC-FD (Iterative Learning

Control-Frequency Domain) needs to have Error and Value Gained rendered in the Frequency Domain.

Error and Value Gained metrics in any domain are intended to compare an optimized to a not optimized configuration for effectiveness of control. In the annotations of these metrics in psychrometric charts, one reflects the Time Domain and the other the Frequency Domain because the underlying models are based on those domains.

AI in Action: The Psychrometric Views of Chiller Efficiency Configurations

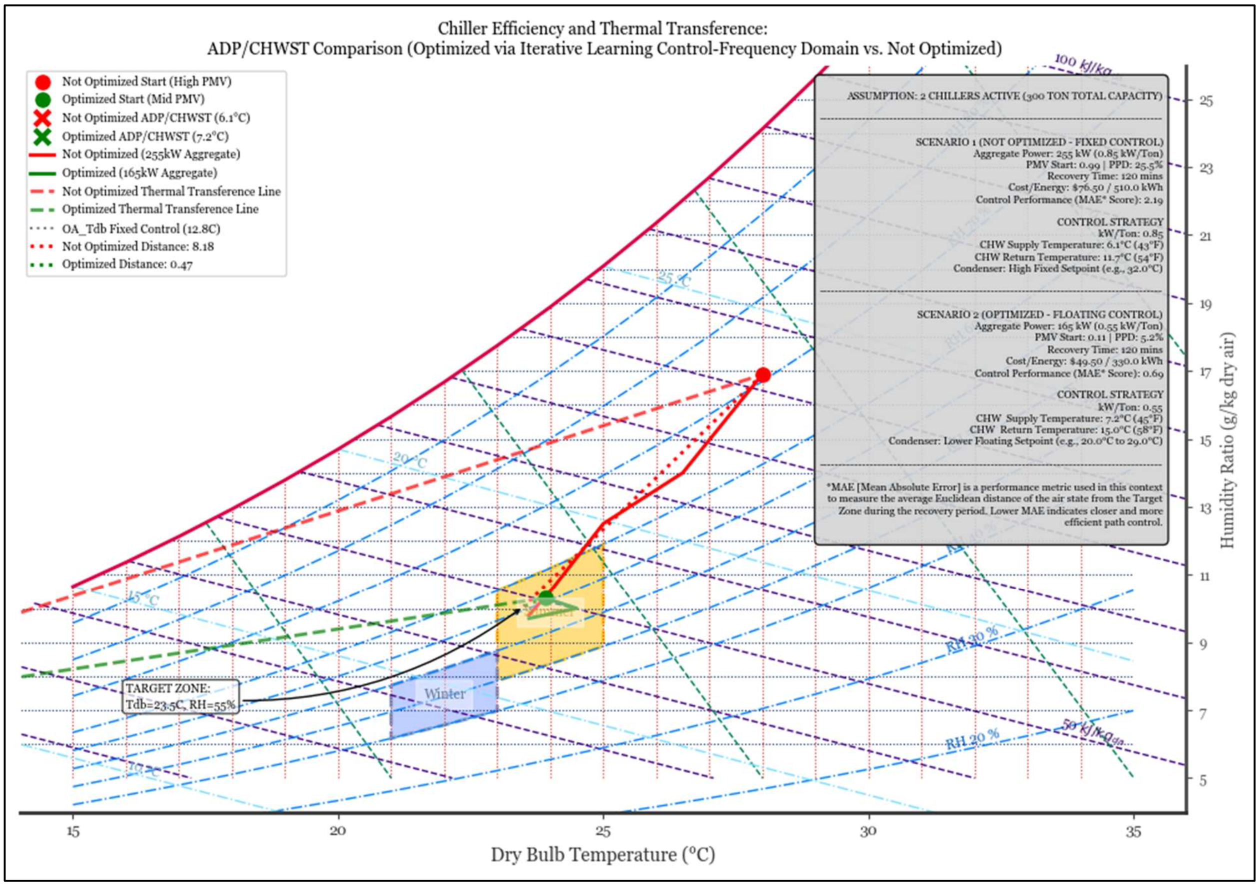

Figure 6: Chiller Psychrometric Chart

Observations

The thermal transference lines for a chiller in **Figure 6** represent the air's final condition for an Optimized and a Not Optimized configuration and is governed by the coil's surface temperature, which is in tight tolerance to the Chilled Water Supply Temperature (CHWST). A good proxy for a coil surface temperature point is to plot the Apparatus Dew Point (ADP), a common technique to represent the thermal interface on a psychrometric chart. If ADP is unavailable as a data point to represent the theoretical saturation temperature of the cooling coil's surface, we can closely approximate ADP by using the CHWST given its tight tolerance to coil surface temperature.

By plotting the two different CHWST values as their theoretical ADPs on the saturation curve, we can represent the thermal transference constraints imposed by the two configurations. It is important to note that the viewing range of the psychrometric chart is limited to temperatures between 15 and 35*C but that the CHWSTs for the Optimized and Not Optimized systems extend to values <15*C, meaning that the ADPs are beyond the left boundary of the chart and the thermal transference lines are only partly displayed. However, it is important to understand what happens just beyond the lower end of the temperature range.

Apparatus Dew Point (ADP): The Thermal Transference Link

Metric	Not Optimized *C	Not Optimized *F	Optimized *C	Optimized *F	Notes
CHW Supply Temp (CHWST ≈ ADP)	6.1	43.0	7.2	45.0	The CHWST sets the lower limit of the thermal sink (proxy for coil surface temperature).
CHW Return Temp (CHWRT)	11.7	53.0	15.0	59.0	The higher the CHWRT, the greater the demonstrated heat absorption capacity or efficiency is.
Delta T	5.6	42.0	7.8	46.0	The thermal transference line is the coil's thermal constraint on the psychrometric chart.

Table 4: Apparatus Dew Point

The CHWST sets the lower limit for how cold the air leaving the coil can be and thus dictates the coil's performance. The ADP is the theoretical point on the psychrometric chart that represents this limit.

The ADP is the theoretical, constant temperature of the cooling coil's surface. It lies on the saturation curve (100% relative humidity line) and serves as the thermal target for the cooling process. When air passes over a cooling coil, its state is driven toward the ADP. The coil's surface temperature - which is very close to the CHWST - is the limiting factor for how much heat can be removed and how much moisture can be condensed. ADP is often represented by a single point on the saturation curve. The line representing the air-cooling process (the solid red and green recovery paths on the psychometric chart) is theoretically directed toward their respective ADP points.

The Optimized system's design in **Table 5** is nearly perfectly matched to the thermal load and moisture requirements of its target starting condition (23.9*C / 55% RH). The 7.2*C CHWST is the highest practical temperature that still provides just enough latent cooling (moisture removal) to maintain the room's humidity ratio (W) without wasting energy by running the chillers colder

Paul Campbell
Vice President of AI & ML Solutions, facil.ai

than necessary. In other words, greater proximity reflects an optimized state with an energy intensity of 0.55 kW/Ton per the annotation in **Figure 4**.

Crossover point proximity to the Optimized ADP and the Role of Slope

Scenario	**P**	**T_{db}**	**W**	**Slope (m)**	**Crossover Point**	
					T_{db}	**W**
Not Optimized	**Coil Inlet (P1)**	28.0	15.0	0.411	**7.9**	**6.7**
	Coil Outlet / ADP (P2)	6.1	6.0			
Optimized	**Coil Inlet (P1)**	23.9	10.0	0.204		
	Coil Outlet / ADP (P2)[1]	**7.2**	**6.6**			

Table 5: Thermal Transference lines

If the Not Optimized system were to receive the same relatively dry and cool air that the Optimized system handles (i.e., air near 23.9*C / 55% RH), the crossover point's lesser proximity to its CHWST indicates:

1. The older configuration would continue to cool and dehumidify the air well past the required point (Latent Cooling Overkill), using its colder 6.1*C CHWST to remove additional moisture that is not needed for comfort.
2. This extra moisture removal requires the chillers to work at greater intensity and is the fundamental source of the energy penalty associated with a higher energy intensity (0.85 kW/Ton) to accommodate a lower CHWST (6.1*C).
3. The shallower slope of the Optimized P1-P2 line reflects efficiency, i.e., more work achieved with less effort as measured in kW/ton [Slope (m) = Latent Heat Change / Sensible Heat Change] than reflected by the Not Optimized P1-P2 line.

In essence, the greater proximity between the crossover and the Optimized ADP and the lesser slope graphically confirms that the Optimized strategy is using an efficient CHWST that provides just enough latent cooling, while the Not Optimized strategy wastes energy on excess latent cooling capacity with no tangible comfort benefit.

Paul Campbell
Vice President of AI & ML Solutions, facil.ai

Control Performance Measurement: RSME and Value Gained

Error:

$$RSME = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_i)^2}$$

Valuation:

$$Log\ Value\ Gained_k = ln\left(1 - \frac{RSME_k}{RSME_{Initial}}\right) + \epsilon$$

Equation 2: Error and Valuation in the Frequency Domain

As for scoring the two illustrated configurations for effectiveness of control, Root Mean Squared Error (RMSE) is a metric used in the frequency domain to quantify the average difference between a set of predicted frequency values (or a model's frequency response) and the actual, observed, or true frequency values. The Value Gained transformation does the following:

- Compresses Large Numbers: An RMSE of 10 is mapped to ln(10), ≈2.3.
- Magnifies Small Numbers: An RMSE of 0.1 is mapped to ln(0.1), ≈-2.3.
- Visualizes Convergence: When you plot the Log RMSE History over k+=1 iterations, the steep initial drop (from 10 to 1) and the long, slow convergence toward zero (from 1 to 0.001) are both clearly represented as a continuous, downward-sloping line, allowing for a proper representation of value gained in the persistence phase.