Abstract—Demand-side control is playing an increasingly important role in smart grid control strategies. Modeling the dynamical behavior of a large population of appliances is especially important to evaluate the effectiveness of various load control strategies. In this paper, a high accuracy aggregated model is first developed for a population of HVAC units. The model efficiently includes statistical information of the population, systematically deals with heterogeneity, and accounts for a second-order effect necessary to accurately capture the transient dynamics in the collective response. Furthermore, the model takes into account the lockout effect of the compressor in order to represent the dynamics of the system under control more accurately. Then, a novel closed loop load control strategy is designed to track a desired demand curve and to ensure a stable and smooth response.

I. INTRODUCTION

Demand-side control (DSM) is considered to be a promising way to resolve many emerging challenges of the power system. Traditional DSM studies mostly focus on peak shaving and load shifting that take place at a relatively slow time scale. With the development of smart grid concepts and communication techniques, the real-time control of a large population of electric loads has received considerable research attention. Among all electric loads, thermostatically controlled loads (TCLs) account for a large fraction of electric demand. HVAC (Heating, Ventilation and air conditioning) systems and water heaters are examples of TCLs. They use local hysteresis control to maintain either air or water temperature within a prescribed band around the temperature set point. Since TCLs allow service interruption for a certain period of time, they are good candidates to provide grid services.

Dynamic modeling of TCLs was first studied in [1] and [2]. In [1], aggregate load models are designed to study the effects of cold load pickup after a service interruption. Functional models of devices, which account for factors such as weather and human behavior, are developed in [2]. A model of a large number of similar devices is then obtained through statistical aggregation of the individual component models. In [3], aggregate models are developed based on solving the coupled Fokker-Planck equations and the need for having a heterogeneous population is motivated. A survey of different load control programs for providing various power system services is presented in [4]. The main contribution of [5] is to develop a finite-dimensional aggregated state space model for air conditioning loads based on solving a bilinear Partial Differential Equation (PDE) model. A statistical model is developed in [6] based on Markov Chains to represent heterogeneity in a population. However, the second-order dynamics caused by solid mass temperature in TCLs are neglected in both detailed and aggregated representations and the uncontrolled aggregated model is not accurate when compared with detailed simulations. In addition, an aggregated state transition based controllable model for a homogeneous population of HVAC units and water heaters is developed in [7] using detailed physical models.

In all the aforementioned approaches, the aggregated models are developed for steady state conditions, i.e., when there is no demand response or time varying changes in population and system configuration (weather conditions, solar gains, heat conductance, etc.). Furthermore, the models ignore heterogeneity in the TCLs population, which is extremely important to capture all the dynamics and thereby making it difficult to design effective control strategies. The approach proposed in [7] is extended in [8] to include simple demand response strategies (thermostat setback) for HVAC units and the results are validated against GridLAB-D, which is an open source distribution system simulation tool [9].

In our recent work [10], an aggregate model for a heterogeneous population of HVACs was developed that is able to capture demand response and time-varying effects of the system. The developed models include second-order dynamics due to the thermal mass of the house, which is ignored in many works dealing with residential buildings. A clustering strategy was also developed to systematically deal with heterogeneity. The developed model can accurately estimate all the transients caused due to demand response, providing a valuable tool for stability analysis of the integrated transmission and distribution system under different demand response strategies.

In this paper, we consider the closed-loop control design of a population of HVAC units under realistic conditions. In practice, a certain amount of time delay between successive "Off" and "On" cycles is required to prevent short cycling of a HVAC system. This compressor time delay effect has not been considered for most aggregate models in the literature. One contribution of this paper is the development of an aggregate model that incorporates the time delay (or the so-
called “lockout”) effect of individual HVAC systems. We also develop a novel closed loop load control strategy for tracking reference demand curves. The proposed load controller is fully responsive and non-disruptive. It achieves the control objective without sacrificing the end-use performance. The proposed aggregated modeling and control strategies are validated through realistic simulations using GridLAB-D. The simulation results indicate that the proposed approach can effectively manage a large number of HVAC systems to provide various demand response services, such as frequency regulation and peak load reduction.

II. INDIVIDUAL HVAC UNIT

This section will discuss some characteristics of a HVAC system. We first introduce a physical model to describe the thermodynamics of a HVAC unit. The compressor time delay effect will also be discussed, which is an important operating constraint of a HVAC unit for demand response studies.

A. Dynamics of Air and Mass Temperatures

Many models have been developed in the literature to describe the thermodynamics of an HVAC unit [11], [12], [13]. In this paper, the popular Equivalent Thermal Parameter (ETP) model [9] is adopted, which describes the dynamics of air and mass temperatures by coupled first-ordered ordinary differential equations:

\[
\begin{align*}
\dot{x}_a &= \frac{1}{C_a}[x_m H_m - x_a(U_a + H_m) + Q_a + T_o U_a] \\
\dot{x}_m &= \frac{1}{\epsilon_m}[H_m (x_a - x_m) + Q_m]
\end{align*}
\]

Here, \(x_a\) is the air temperature and \(x_m\) represents the inner mass temperature. To simplify notation, we collect all the model parameters to form a parameter vector \(\theta\) as follows:

\[
\theta = [U_a, T_o, H_m, C_a, C_m, Q_i, Q_s, Q_h]^T,
\]

where \(U_a\) is the conductance of the building envelope, \(T_o\) is the outdoor air temperature, \(H_m\) is the conductance between the inner air and inner solid mass, \(C_a\) is the thermal mass of the air, \(C_m\) is the thermal mass of the building materials and furnishings, \(Q_a\) is the heat flux into the interior air mass and \(Q_m\) is the heat flux to the interior solid mass. The total heat flux \(Q_a\) consists of three main contributing factors, \(Q_i\), \(Q_s\) and \(Q_h\), where \(Q_i\) is the heat gain from the internal load, \(Q_s\) is the solar heat gain, and \(Q_h\) is the heat gain from the heating/cooling system. Depending on the power state of the unit, the heat flux \(Q_a\) could take the following two values:

\[
Q_a^{on} = Q_i + Q_s + Q_h \quad \text{and} \quad Q_a^{off} = Q_i + Q_s
\]

The readers are referred to [9] for the details of the relation between the ETP model parameters and the physical properties of the house. The ETP model can be written in the state-space form by defining the state vector as \(x = [x_a, x_m]^T\):

\[
\dot{x}(t) = Ax(t) + B_{on/off}.
\]

For a given control deadband \([u_{set} - \delta/2, u_{set} + \delta/2]\) with setpoint \(u_{set}\) and deadband size \(\delta\), the HVAC unit switches on and off when the air temperature hits the upper and lower boundary of the deadband. A typical state trajectory is illustrated in Fig. 1. The air temperature dynamics are different for each cycle, especially for the first cycle when the mass temperature is far lower. A detailed discussion on the effect of the mass temperature can be found in [10], [14].

B. Compressor Minimum Off-time

When the compressor of a HVAC system is turned “off”, the air pressure in the chamber is high and a certain amount of time is needed for the pressure to even out. Restarting the compressor under pressure may cause physical damage [15]. Compressor time delay relays are typically installed for HVAC units that are controlled by electronic or programmable thermostats. These relays are used to keep the compressors in the “Off” state for a minimum off-time, for instance, five minutes. During the minimum-off time period, the HVAC unit is “locked”, and any switching-on control signal will be ignored. The dashed lines in Fig. 1 illustrate the part of the state trajectory during which the unit is “locked” due to the spontaneous switching-off at the lower boundary of the control deadband. This property of HVAC systems will be referred to as the “lockout” effect in the rest of this paper. Any practical demand response study involving HVAC systems should consider the lockout effect.

III. REALISTIC CONTROL OF HVAC AGGREGATION

This section will develop a control strategy of the aggregate HVAC units with consideration of the lockout effect and other realistic conditions.

A. Aggregation of HVAC Systems

As the population size increases, it becomes more challenging to model the detailed state trajectories of all the HVAC systems. To address this issue, many aggregate models have been developed in the literature [5], [6], [8]. The basic idea of all these aggregate models is to characterize the time-course evolution of the temperature density of the population, instead of considering all the individual temperature trajectories. In discrete time, the aggregate models can
be represented in the following form:

\[ q(k + 1) = Gq(k) \]  

(1)

where \( q(k) \triangleq [(q^{on}(k))^T, (q^{off}(k))^T]^T \), is the density vector of size \( 2n \), where \( n \) is the number of temperature bins. Each entry of \( q^{on}(k) \) (or \( q^{off}(k) \)) represents the fraction of loads whose mode is “On” (or “Off”) and whose air and mass temperatures are in the corresponding temperature bin. The models in the literature differ only in their ways of determining the state transition matrix \( G \). The output of the model is the total power consumption of the population, which depends linearly of the density vector:

\[ y(k) = [1, \ldots, 1, 0, \ldots, 0]q(k) \triangleq Cq(k) \]  

(2)

This paper will use the aggregate model developed in our previous work [10], which is obtained based on the 2nd-order ETP model of an HVAC unit. To obtain the aggregate model, a two dimensional grid on \( x_a \) and \( x_m \) is mapped into the density vector \( q \in \mathbb{R}^{2n_a \times n_m} \), where \( n_a \) and \( n_m \) are the numbers of temperature bins in \( x_a \) and \( x_m \), respectively. When the HVAC loads are homogeneous, they all share the same parameter vector \( \theta \). A first-principle based algorithm is proposed in [10] to compute the state transition matrix \( G \). In addition, to account for the heterogeneity among real HVAC populations, we cluster all the ETP parameters using the “kmeans” function in MATLAB. The center of each cluster is used as a common parameter value to compute the homogeneous state transition matrix for each cluster. The total aggregate response is obtained by the weighted sum of the outputs of all the clusters. We have shown in [10] that the aggregate model can accurately capture both the transient and the steady-state collective dynamics of the HVAC systems.

B. Existing Aggregated Control Approaches

Several model-based controllers of aggregated HVAC systems have been proposed in the literature [5], [6], [16]. One common control approach is the so-called thermostat setback program. Under this approach, a central controller broadcasts universal set point changes to all the HVAC units. A small change in set point of a large number of HVAC units can result in a significant change in the aggregated power. However, right after the change, a large undesirable rebound is often observed. The problem has been studied in [16], where some mathematical approximations of the transient dynamics are made. Under the assumption that the aggregate power can be measured and transmitted to the controller every few seconds, a closed-loop controller is developed that broadcasts the set point change signal to the population. However, the approach does not consider the “lockout” effect, thereby limiting its practical application.

Another widely used aggregate control approach is the so-called “toggling” control [6]. Under this approach, a central controller broadcasts a vector signal \( u \) to all the devices to directly control their “On/Off” power modes. Each bin in \( u \) has a probability of switching in a certain range of temperature. For example, when \( u_i \in [0, 1] \), the HVAC units in the “Off” state with the temperatures in the \( i^{th} \) bin will have the probability of \( |u_i| \) to turn “On”. Similarly, a control value \( u_i \in [-1, 0] \) will turn off the corresponding loads probabilistically. The temperature difference between the neighboring bins in \( u \) is about one or two orders smaller than the resolution of thermostats, which is typically around \( 1^\circ F \). In this case, the information of the probability of switching contained in \( u \) becomes unrecognizable at the device level. Another issue of the controller is that it also does not consider the “lockout” effect.

Both types of control strategies mentioned above control the aggregate load of HVAC units effectively under the assumption that every HVAC unit can respond to control signals in a short time period, for instance, several seconds. However, the units may not be able to respond due to the “lockout” effect. When a HVAC unit is “locked”, it cannot be switched, even if the temperature is beyond the deadband when the set point is changed by the control signal. Similarly, the “locked” unit cannot be toggled by the broadcasted vector \( u \). The effect restricts the possibility of controlling in that short period. If the “locked” populations are included in the aggregate model and kept from being switched, it would make the control of HVAC units in a short time period possible. This issue will be addressed in the following subsection.

C. Proposed Aggregated Control Strategy

In this subsection, we will develop an aggregated control strategy, which explicitly considers the “lockout” effect of HVAC units. In addition, the developed control strategy is temperature independent, that is, it does not require the temperature information to generate the control signal. Under our proposed control strategy, the system operator will broadcast a scalar control signal \( \alpha(k) \in [-1, 1] \). Each HVAC unit will interpret the control signal probabilistically based on its own power mode. For example, if \( \alpha(k) > 0 \), then each “Off” unit will have \( \alpha(k) \) probability of turning “On” right away. On the other hand, if \( \alpha(k) < 0 \), each “On” unit will turn off with probability \( |\alpha(k)| \).

The evolution of the state vector \( q(k) \) under the above control strategy can be described by the following modified aggregate model:

\[ q(k + 1) = \begin{cases} GB_1(\alpha(k))q(k), & \text{if } \alpha(k) \geq 0 \\ GB_2(\alpha(k))q(k), & \text{if } \alpha(k) < 0 \end{cases} \]  

(3)

where \( B_1(\cdot) \) and \( B_2(\cdot) \) are matrix-valued functions of the control \( \alpha(k) \) defined as follows:

\[ B_1(\alpha) = \begin{bmatrix} I_n & \alpha I_n \\ 0 & (1-\alpha)I_n \end{bmatrix}, \quad B_2(\alpha) = \begin{bmatrix} (1+\alpha)I_n & 0 \\ -\alpha I_n & I_n \end{bmatrix}, \]

for an arbitrary control value \( \alpha \in [-1, 1] \). Here, \( I_n \) denotes the identify matrix of dimension \( n \). Equation (3) describes the density dynamics under the proposed control strategy \( \alpha(k) \) without considering the lockout effect. As described in Section II-B, the “locked” populations cannot be turned “on” until the end of the minimum off-time. To account for this restriction, we introduce another state vector \( q^L = \ldots \)
\[ q^L = \begin{bmatrix} q^L_1 \\ \vdots \\ q^L_{n_L} \end{bmatrix} \] whose dimension \( n_L = \frac{\tau}{\Delta t} \), where \( \Delta t \) is the discrete time unit. The \( i \)-th entry of \( q^L \) is the amount of locked population that will be released after \( n_L - i + 1 \) discrete time units. The total locked population is given by \( y^L = G^L q^L \), where \( G^L \) is a row vector of dimension \( n_L \) with all entries equal to 1. Notice that the lockout population is a subset of the “Off” population and cannot be switched “On”. At any time \( k \), the actual population that can be freely switched “On” is given by:

\[
\hat{q}(k) = \begin{bmatrix} I_n & 0 \\ 0 & I_n(1 - \frac{q^L(k)}{Q_{off}(k)}) \end{bmatrix} q(k) \triangleq E(y^L(k))q(k) \quad (4)
\]

where \( Q_{off} \) is the summation of all the populations in “Off” state. It is beneficial to think of the entire evolution during each discrete time period as two sequential stages, where during the first stage the population switches according to the control signal \( \alpha(k) \), while during the second stage the population evolves naturally according to the state-transition matrix \( G \). Denote \( q^+(k) \) as the density vector after the first stage of time step \( k \). Then, with consideration of the lockout effect, the modified aggregate model is given by:

\[
\begin{cases}
q^+(k) &= B_1(\alpha(k))\hat{q}(k) + (q(k) - \hat{q}(k)) \\
q(k + 1) &= Gq^+(k) \\
q^L(k + 1) &= G^L q^L(k) + B^L \cdot S \cdot q^+(k),
\end{cases} \quad (5)
\]

\[
\begin{cases}
q^+(k) &= B_2(\alpha(k))q(k) \\
q(k + 1) &= Gq^+(k) \\
q^L(k + 1) &= G^L q^L(k) + B^L \cdot (\|\alpha\| \cdot y(k) + S \cdot q^+(k)),
\end{cases} \quad (6)
\]

where \( B^L = \begin{bmatrix} 1, 0, \ldots, 0 \end{bmatrix}^T \) and \( S \) is a \( 1 \times 2n \) matrix such that \( S \cdot q^+ \) is the total amount of loads that will turn “off” and become locked if the system evolves autonomously (without additional control) from \( q^+ \) during the second stage of time step \( k \). The matrix \( S \) can be derived as follow:

\[
S = \begin{bmatrix} [0, \ldots, 0, 1, \ldots, 1]^T \end{bmatrix} G \begin{bmatrix} I_n \\ 0 \\ 0 \end{bmatrix}
\]

The \( G^L \) matrix in equations (5) and (6) is an \( n_L \times n_L \) matrix that determines the evolution of the lock-out population. It is given by:

\[
G^L = \begin{bmatrix}
0 & 0 & \cdots & \cdots & 0 \\
1 & 0 & \ddots & \ddots & \vdots \\
0 & 1 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 & 0
\end{bmatrix}
\]

Fig. 2 illustrates the state transition of the modified aggregate model. In the proposed model, unlike other aggregate models without considering lockout effect, the populations in the “On” state never flow directly into “Off” state. Instead, the population first flows into \( q^L \), when hitting the lower boundary of the deadband or being toggled by the control signal. After evolving to \( q^L \), the population has been “locked” for at least the minimum off-time \( \tau \), and is thus released to the unlocked “Off” state. One feature is that even if every \( q^L \) is a scalar, the populations in \( q^L \) still evolves according to the state-transition matrix \( G \). As a result, when the populations are released from the “locked” states by \( G^L q^L \), each of them flows back to the bin corresponding to its temperature and “On/Off” state. The cycle completes once the population returns back to \( q_{on} \) when reaching the upper boundary of the deadband or being toggled “On” by the control signal \( \alpha \).

The control signal \( \alpha \) is designed to match the real density of “on” state \( y_{real} \) for a given desired trajectory \( y_{ref} \). Since \( y_{model}(k + 1) = Cq(k + 1) \) is considered to be a good estimation of \( y_{real}(k + 1) \), we use the MATLAB function “fzero” in equation (5) and (6) to find the \( \alpha(k) \) such that \( y_{ref}(k + 1) - y_{model}(k + 1) \approx 0 \). The \( y_{ref}(k + 1) \) is then expected to follow the \( y_{ref}(k + 1) \) as \( y_{model}(k + 1) \) does.

The total amount of “locked” populations, considered in the modified aggregate model are kept from switching by the control signal \( \alpha \). This design allows each HVAC unit to ignore the commands of turning “on” before the end of the minimum off-time. Therefore, the central controller is able to send control signals to the device level in a short period without concerns of forcing some populations to start overloading. This prevents an inaccurate estimation of the aggregate load due to the “locked” populations, which are not turned “on”. One issue is that if too many devices are “locked”, the number of available controllable units may be few. The problem would be resolved when a large enough population size of HVAC is available.

IV. Simulation Results

In this section, simulation studies were performed to illustrate the benefits of the proposed aggregated control strategy. 5000 sets of physical parameters are generated, which are
randomly distributed around their nominal values with a certain amount of variance. Each of them represents the real condition in one house. Based on the relationship between the physical parameters and the ETP model described in [9], 5000 sets of ETP model parameters are obtained and used in all the simulations in this section. The proposed aggregate model with lockout consideration is validated in Section IV-A. The proposed aggregated control strategy is then tested against different scenarios of the possible demand response service including regulation and load reduction. In those case studies, HVAC units are assumed to consume 5(kW) on average.

A. Aggregate Response with Lockout Effect

The lockout effect plays an important role when HVAC units are subjected to the control signal $\alpha$ described in III-C. The simulation in this subsection compares the performance of predicting the real aggregate output between the modified aggregate model and the original one with no consideration of “locked” populations. The real aggregate output is obtained by simulating each HVAC unit with its ETP model. As shown in Fig. 3, the modified model matches the real aggregate load better than the original one, especially when $\alpha$ is positive. The original model overestimates the “On” populations when $\alpha$ is positive because some “locked” populations are assumed to be turned “on”, while this is actually not the case. On the contrary, the modified aggregate model estimates the “locked” populations and aggregated power accurately.

B. Scenario 1: Regulation Service

Since HVAC units can be turned on and off very quickly if they are not subject to lockout, they have the potential to provide fast response as desired by regulation services. In this subsection, we assume 5000 HVAC units are incorporated to provide the regulation service in PJM, which is one of the main regional transmission organization (RTO) in North America. Regulation is a service which manages a continuous balance of electricity demand and loads in response to the changes in electricity use. In PJM, the total required regulation is roughly one percent of the regional load. The readers are referred to [17], [18] for the details of regulation service.

The 5000 HVAC units are collected to provided up-regulation and down-regulation with capability of 2.5(MW). When the set point of those populations is $75^\circ F$, the average aggregate power consumption is around 3.5(MW). Therefore, 3.5(MW) is taken as a baseline of the 5000 HVAC units in the following simulation. The details of the methods on estimating the baseline load are in [19]. With the baseline given, the HVAC units are then assigned to follow a dynamic regulation signal, which is a test signal downloaded from the PJM website [20]. The signal was adjusted such that the regulation range is 2.5(MW). During the regulation period, each HVAC unit receives control signals $\alpha$ and sends the measured power consumption data to central controller every fifteen seconds. Fig. 4 shows that the controlled aggregated loads can follow the reference signal very accurately. In the simulation, each load is assumed to have a 5-minute compressor time delay.

C. Scenario 2: Direct Load Control

Direct load control (DLC) is a demand-side management program that curtails the power consumption of consumers during peak demand period. The same 5000 HVAC units are assumed to participate in a DLC program. Consider the scenario where 50% of them are switched “off” directly after 1 hr. in Fig. 5 and kept in “off” state until 30 minutes after. By the end of the load reduction period, they are released simultaneously at 1.5 hr. The effect of cold load pickup is then observed, which may damage the distribution system. We address this issue by controlling the load based on the modified aggregate model for not only the load reduction period, but also thirty minutes afterward. During the load reduction, aggregate load of HVAC units is controlled at 1.7 (MW), instead of the steady state 3.5 (MW). For the following 30 minutes after the end of the load reduction, the load is controlled at 4.2 (MW). The value is slightly higher than the steady state for 30 minutes to provide a room to spread out the turning “On” time of those “Off” units from the load reduction. As a result, the “Off” populations are kept from turning “On” simultaneously and a smooth aggregate
response is obtained. Another benefit of the second control strategy is that the temperature of every user is kept between the dead-band. The first DR approach made those devices turned “Off” for at least thirty minutes, which will cause too much discomfort to the users.

V. CONCLUSION

We have developed an aggregate model of a population of HVAC units under realistic conditions. We considered the lockout effect of the compressor in order to identify the number of HVAC units in the population that are not controllable subject to the load control signal. It is the first time that the lockout effect is considered in the aggregate modeling of HVAC units. Using the developed aggregate model, we designed a novel closed-loop load control strategy for the population of HVAC units to track a prescribed demand curve without affecting the end-use performance. The simulation results demonstrated the effectiveness of the proposed load control strategy.

REFERENCES