

Inference of Building Occupancy Signals Using Moving Horizon Estimation and Fourier Regularization*

Victor M. Zavala

Mathematics and Computer Science Division

Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA

Abstract

We study the problem of estimating time-varying occupancy and ambient air flow signals using noisy carbon dioxide and flow sensor measurements. A regularized moving horizon estimation formulation is proposed that forces time-varying signals to fit smooth Fourier expansions. We demonstrate that the regularization approach makes the estimator robust to high levels of noise. In addition, it requires minimal information about the shape of the signals. Computational experiments using synthetic and real data are presented to demonstrate the effectiveness of the approach.

Keywords: occupancy, moving horizon estimation, carbon dioxide, air flow, regularization, Fourier.

1 Introduction

Occupancy estimation is key for control and measurement and verification (M&V) tasks. From a control standpoint, such information is needed in order to determine the heating and cooling loads as well as ventilation rates necessary to maintain comfort and air quality conditions [21, 12]. This is particularly important in predictive control systems where load predictions (i.e., disturbances) are necessary. Occupancy estimates are required in M&V tasks in order to construct baseline building energy use models [6]. These baseline models are later used to assess energy savings of retrofits (including new control systems). In both these tasks, low-resolution and building-wide estimates are often sufficient [27, 14, 12]. Low-resolution estimates are also important because both control and M&V systems need to be low cost. Consequently, the use of a small number of carbon dioxide and air flow sensors is preferred. More advanced and sophisticated applications of occupancy information include emergency response and occupancy behavior analysis for building design. These tasks require high-accuracy and zone-level estimates [7, 13, 25, 23]. Consequently, the required sensor networks tend to be more expensive because they involve vision and radio frequency identification (RFID) systems, among others. In this paper, we focus on occupancy estimation using carbon dioxide and air flow measurements.

*Preprint MCS/ANL-P4031-0213.

The literature on occupancy estimation techniques abounds. To the best of our knowledge, however, constrained moving horizon estimation (MHE) has been considered only in [13]. The rest of the approaches available are based on simplified maximum likelihood formulations and steady-state model approximations [25, 14, 27, 8]. The benefits of MHE are well known [5, 32, 9]. Of particular interest is the fact that MHE provides a flexible framework to *regularize* estimates by using priors and constraints derived from diverse sources of information and engineering intuition. In addition, MHE can handle nonlinearities in a systematic manner, and it can be adapted to simultaneously estimate time-varying input and parameter signals (e.g., errors-in-variables formulations). These practical benefits have been demonstrated in [4, 13, 32]. Stability properties and extensions to hybrid systems have been presented in [19, 1, 31, 3]. Techniques to estimate noise statistics have been presented in [15, 17, 11].

In this work, we explore a different form of regularization for MHE. We propose to regularize time-varying signals by forcing them to fit smooth Fourier expansions. The need for regularization is motivated by the occupancy estimation problem where occupancy and air flow signals tend to be highly sensitive to sensor noise [27]. In other words, the signals tend to be weakly observable. In addition, limited information about the shape of the occupancy trends is normally available to enable prior and constraint regularization. The reason is that occupancy trends strongly depend on the building type. For instance, office buildings and schools have entirely different profiles [8, 27, 2, 20]. The problem of estimating time-varying input signals is also important in industrial systems where trends of fouling coefficients and kinetic parameters are needed [32, 5, 22]. Other application areas are disease infection models [29] and psychology [18].

The idea of regularizing time-varying signals using by smoothing functions has been widely used in regression analysis [28, 18, 16]. These studies use splines and piece-wise polynomials with known derivative and growth behavior in order to induce desired signal behavior. We exploit similar capabilities within general MHE formulations to infer time-varying occupancy and air flow signals. We use Fourier expansions to capture the inherent periodicity of occupancy and flow signals prevailing in buildings. We demonstrate that MHE formulations regularized in this form can significantly reduce estimate volatility and sustain strong levels of noise.

The paper is structured as follows. In Section 2 we describe the building system model. In Section 3 we present the MHE formulation under study and describe the Fourier regularization procedure. A series of numerical experiments is presented in Section 4 using synthetic and real data. In Section 5 we summarize our conclusions and briefly discuss future work.

2 Occupancy Estimation

We consider the problem of inferring the number of occupants $n_{oc}(\tau)$ in a space of volume V using sensor signals of carbon dioxide $C(\tau)$ and inlet air flow $q_{in}(\tau)$. The building-wide mass balance of

the system is given by

$$\frac{1}{\rho(\tau)} \frac{dm}{dt}(\tau) = q_{in}(\tau) - q_{out}(\tau) + q_{inf}(\tau) - q_{exf}(\tau) \quad (2.1a)$$

$$V \frac{dC}{dt}(\tau) = q_{in}(\tau) \cdot C_{in}(\tau) - q_{out}(\tau) \cdot C(\tau) + q_{inf}(\tau) \cdot C_{in}(\tau) - q_{exf} \cdot C(\tau) + G \cdot n_{oc}(\tau). \quad (2.1b)$$

The total air mass is given by $m(\tau)$. The inlet concentration $C_{in}(\tau)$ is the atmospheric concentration and is assumed to be constant. The outlet, infiltration, and exfiltration flows are denoted by $q_{out}(\tau)$, $q_{inf}(\tau)$ and $q_{exf}(\tau)$, respectively. G is the average CO₂ generation rate per occupant, and $\rho(\tau)$ is the air density.

An internal feedback controller ensures that system pressure remains constant, which implies that the mass $m(\tau)$ is constant and $0 = q_{in}(\tau) - q_{out}(\tau) + q_{inf}(\tau) - q_{exf}(\tau)$. Using this relationship, we have that

$$V \frac{dC}{dt}(\tau) = (q_{in}(\tau) + q_{inf}(\tau)) \cdot (C_{in}(\tau) - C(\tau)) + G \cdot n_{oc}(\tau). \quad (2.2)$$

Note that the exfiltration and outlet flows do not appear in the balance equation. Buildings are typically operated under positive pressure (internal building pressure higher than atmospheric) [26, 14]. This effectively minimizes infiltration. Hence, in this case, one can assume that $q_{inf}(\cdot) \approx 0$. Note that this assumption does not hold in the case of natural ventilation [20]. With this, we have

$$V \frac{dC}{dt}(\tau) = q_{in}(\tau) \cdot (C_{in}(\tau) - C(\tau)) + G \cdot n_{oc}(\tau). \quad (2.3)$$

In this balance we assume that sensor measurements for $C(\tau)$ and $q_{in}(\tau)$ are available while the unknown signal is $n_{oc}(\tau)$. We treat $q_{in}(\tau)$ as an unknown signal, however, in order to account for highly noisy signals. This gives an errors-in-variables formulation.

In Section 4 we present evidence that the proposed model is sufficient to obtain consistent building-wide occupancy estimates. The model has been used in other real occupancy estimation studies reported in the literature [12, 27], but it can certainly be extended in a number of ways by including models of infiltration and exfiltration rates and by modeling the recycle ducts [27]. These more detailed models, however, will also require of additional sensor information that might not be available in a typical building. While these model extensions can potentially improve estimates, we prefer to leave them as a topic of future work and instead focus on the MHE formulation and explore the benefits of regularization.

3 MHE Formulation

The sensor signals are assumed to follow Gaussian distributions $C(\tau) \sim \mathcal{N}(\bar{C}(\tau), \sigma_C^2)$, $q_{in}(\tau) \sim \mathcal{N}(\bar{q}_{in}(\tau), \sigma_{q_{in}}^2)$ with known and constant variances $\sigma_{q_{in}}^2, \sigma_C^2$. The means are the sensor readings $\bar{C}(\tau), \bar{q}_{in}(\tau)$. The prior initial condition at time $t - T$, where T is the estimation horizon is also assumed to be Gaussian $C(t - T) \sim \mathcal{N}(\bar{C}(t - T), \sigma_C^2)$ as this is also measured.

3.1 Cost and Constraints

The estimator minimizes the following log-likelihood function at time t :

$$\varphi_{t-T}^t = \int_{t-T}^t \left(\frac{1}{2\sigma_{q_{in}}^2} (q_{in}(\tau) - \bar{q}_{in}(\tau))^2 + \frac{1}{2\sigma_C^2} (C(\tau) - \bar{C}(\tau))^2 \right) d\tau. \quad (3.4)$$

We constrain the estimator to the physical CO₂ dynamics (2.3) defined over $\tau \in [t - T, t]$.

It is well known that bound constraints are an effective way to regularize MHE problems. This helps MHE filter out regions of high probability with no physical meaning (e.g., negative concentrations, unrealistic occupancy levels) [4]. We add the following constraints to the MHE problem

$$L_C(\tau) \leq C(\tau) \leq U_C(\tau) \quad (3.5a)$$

$$L_{n_{oc}}(\tau) \leq n_{oc}(\tau) \leq U_{n_{oc}}(\tau) \quad (3.5b)$$

$$L_{q_{in}}(\tau) \leq q_{in}(\tau) \leq U_{q_{in}}(\tau), \quad (3.5c)$$

for $\tau \in [t - T, t]$. Here, $L_i(\tau), U_i(\tau)$ are lower- and upper-bound trajectories with $i \in \{C, n_{oc}, q_{in}\}$.

3.2 Fourier Regularization

In certain applications, constraints and likelihood priors are insufficient to achieve stable estimates. The reason may be the inherent weak observability of the system [10]. As we will see in Section 4, this is the case for occupancy estimation. We propose to regularize the problem by assuming that occupancy and flow signals vary smoothly in time. Consequently, these can be forced to fit smooth functions, and we can estimate the coefficients as part of the estimation procedure. Because the signals arising in buildings are periodic, we propose to use Fourier expansions. The expansions for occupancy and input flow are given by

$$n_{oc}(\tau) = \sum_{j=0}^{n_{n_{oc}}} (a_{n_{oc}}^j \cdot \cos(j\theta(\tau)) + b_{n_{oc}}^j \cdot \sin(j\theta(\tau))) \quad (3.6a)$$

$$q_{in}(\tau) = \sum_{j=0}^{n_{q_{in}}} (a_{q_{in}}^j \cdot \cos(j\theta(\tau)) + b_{q_{in}}^j \cdot \sin(j\theta(\tau))). \quad (3.6b)$$

These are defined over $\tau \in [t - T, t]$. The expansion domain is symmetrized by using the change of variables,

$$\theta(\tau) = \left(\frac{\tau - (t - T)}{T} \right) \pi - \left(1 - \frac{\tau - (t - T)}{T} \right) \pi. \quad (3.7)$$

This ensures that $-\pi \leq \theta(\tau) \leq +\pi$ with $\theta(t - T) = -\pi$ and $\theta(t) = +\pi$. The number of expansion terms for each signal is given by n_i^j for $i \in \{n_{oc}, q_{in}\}$ with a_i^j, b_i^j , $j = 0, \dots, n_i$, $i \in \{n_{oc}, q_{in}\}$ being the corresponding coefficients.

We highlight that it is in principle possible to tackle the occupancy inference problem by using recent approaches for covariance estimation [17, 10]. In particular, these approaches would treat the occupancy signal as a stochastic process for which its autocorrelation structure (i.e., its dynamics) would be inferred from data. We have not tested these approaches because they are currently

based on unconstrained Kalman estimators. In our case, we are seeking approaches to infer time-varying signals under a general optimization-based MHE framework capable of: i) simultaneously estimate input and disturbance signals, ii) it handles constraints, and ii) it can deal with complex physical models for which Kalman estimators are cumbersome to implement or are computationally intractable [32]. We acknowledge, however, that it would be of interest to compare the performance of both approaches or consider hybrid variants. We leave this as a topic of future work.

3.3 Implementation and Computational Issues

To solve the MHE problem, we discretize the problem applying an Euler scheme with $N + 1$ terms each of length Δt . We assume that the discretization mesh and measurement sampling intervals coincide. After discretization the MHE problem becomes a nonlinear program of the form

$$\min \sum_{k=0}^N \left(\frac{1}{2\sigma_{q_{in}}^2} (q_{in}(\tau_k) - \bar{q}_{in}(\tau_k))^2 + \frac{1}{2\sigma_C^2} (C(\tau_k) - \bar{C}(\tau_k))^2 \right) \quad (3.8a)$$

$$\text{s.t.} \quad (3.8b)$$

$$V \frac{C(\tau_{k+1}) - C(\tau_k)}{\Delta t} = q_{in}(\tau_{k+1}) \cdot (C_{in}(\tau_k) - C(\tau_k)) + G \cdot n_{oc}(\tau_k) \quad (3.8c)$$

$$n_{oc}(\tau_k) = \sum_{j=0}^{n_{noc}} (a_{noc}^j \cdot \cos(j\theta(\tau_k)) + b_{noc}^j \cdot \sin(j\theta(\tau_k))) \quad (3.8d)$$

$$q_{in}(\tau_k) = \sum_{j=0}^{n_{qin}} (a_{qin}^j \cdot \cos(j\theta(\tau_k)) + b_{qin}^j \cdot \sin(j\theta(\tau_k))) \quad (3.8e)$$

$$L_C(\tau_k) \leq C(\tau_k) \leq U_C(\tau_k) \quad (3.8f)$$

$$L_{q_{in}}(\tau_k) \leq q_{in}(\tau_k) \leq U_{q_{in}}(\tau_k) \quad (3.8g)$$

$$L_{n_{oc}}(\tau_k) \leq n_{oc}(\tau_k) \leq U_{n_{oc}}(\tau_k), \quad (3.8h)$$

with $k = 0, \dots, N - 1$. We note that the Fourier expansion constraints induce a regularization effect. When a single expansion term is used, we have a constant signal (estimator is fully constrained) and given by a_i^0 because $\cos(0) = 1, \sin(0) = 0, k = 0, \dots, N$. This gives the maximum possible cost. When the number of expansion terms is equal to the number of discretization steps, we have that the signal at each step can take any value desired. This gives maximum flexibility and the minimum possible cost. In other words, the solution is equivalent to that of the unregularized MHE problem,

$$\min \sum_{k=0}^N \left(\frac{1}{2\sigma_{q_{in}}^2} (q_{in}(\tau_k) - \bar{q}_{in}(\tau_k))^2 + \frac{1}{2\sigma_C^2} (C(\tau_k) - \bar{C}(\tau_k))^2 \right) \quad (3.9a)$$

$$\text{s.t.} \quad (3.9b)$$

$$V \frac{C(\tau_{k+1}) - C(\tau_k)}{\Delta t} = q_{in}(\tau_{k+1}) \cdot (C_{in}(\tau_k) - C(\tau_k)) + G \cdot n_{oc}(\tau_k) \quad (3.9c)$$

$$L_C(\tau_k) \leq C(\tau_k) \leq U_C(\tau_k) \quad (3.9d)$$

$$L_{q_{in}}(\tau_k) \leq q_{in}(\tau_k) \leq U_{q_{in}}(\tau_k) \quad (3.9e)$$

$$L_{n_{oc}}(\tau_k) \leq n_{oc}(\tau_k) \leq U_{n_{oc}}(\tau_k). \quad (3.9f)$$

We note that the expansion constraints (3.8d) and (3.8e) can be written in matrix form as

$$\mathbf{n}_{oc} = \Pi_{n_{oc}} \mathbf{a}_{n_{oc}} + \Gamma_{n_{oc}} \mathbf{b}_{n_{oc}} \quad (3.10a)$$

$$\mathbf{q}_{in} = \Pi_{q_{in}} \mathbf{a}_{q_{in}} + \Gamma_{q_{in}} \mathbf{b}_{q_{in}}. \quad (3.10b)$$

Here, \mathbf{n}_{oc} , \mathbf{q}_{in} , \mathbf{a}_i , and \mathbf{b}_i are variable vectors for the signals and coefficients, respectively. The matrices Π_i have (k, j) entries given by $\cos(j \cdot \theta(\tau_k))$, $k = 0, \dots, N$, $j = 0, \dots, n_i$, $i \in \{n_{oc}, q_{in}\}$. A similar definition follows for matrices Γ_i with entries $\sin(j \cdot \theta(\tau_k))$. We emphasize that these are dense $(N + 1) \times (n_i)$ matrices and induce dense blocks in the Karush-Kuhn-Tucker (KKT) system of problem (3.8). This will deteriorate the computational efficiency of sparse factorization routines used in optimization solvers because they limit pivoting flexibility and induce large amounts of fill-in, as we demonstrate in the following section. This situation can be managed, however, by reducing the number of expansion terms and the horizon length. In particular, in the next section we will see that as the horizon is shortened, the number expansion terms can also be reduced, leading to significant improvements in computational time.

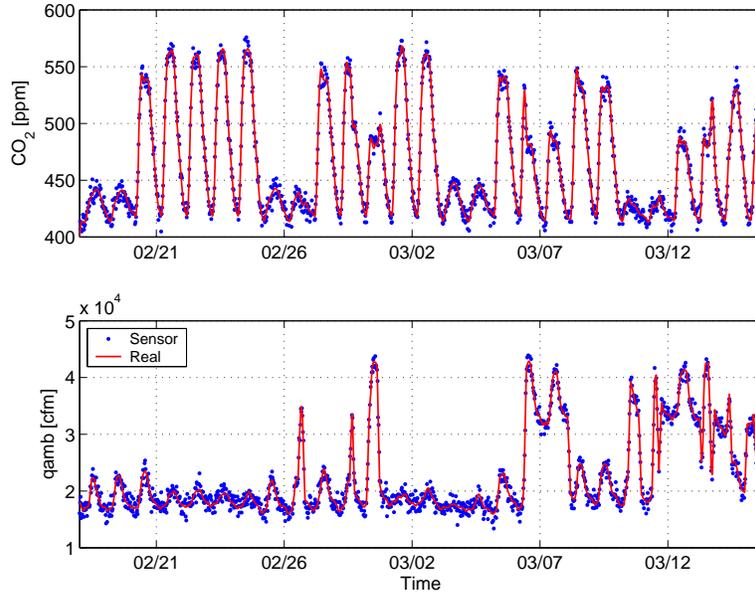


Figure 1: Real and sensor signals for CO₂ and inlet flow used in synthetic study.

4 Numerical Studies

We demonstrate the effectiveness of the regularization procedure using a study with *synthetic* data. We use a study with *real* sensor data to demonstrate that the model coupled to regularized MHE provide realistic occupancy estimates.

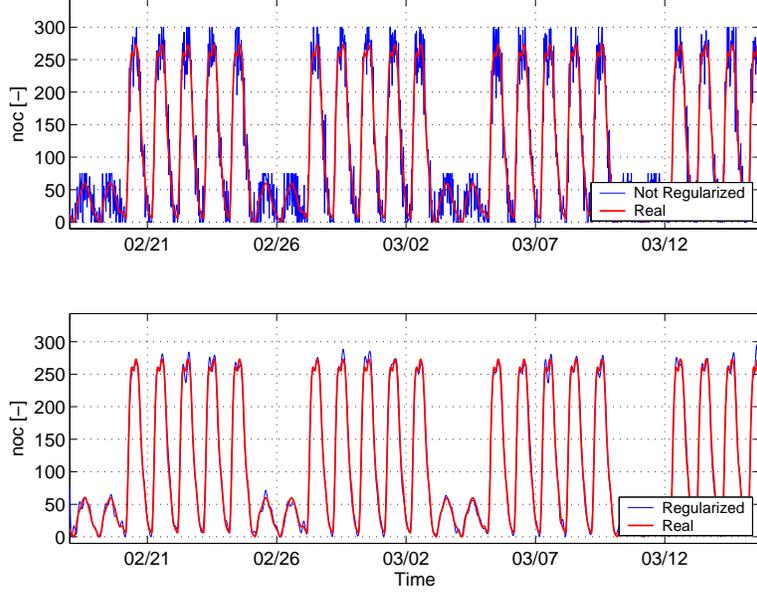


Figure 2: Real and inferred occupancy signals for MHE under no regularization (top). Real and inferred signals for MHE with regularization (bottom).

4.1 Synthetic Data

For the synthetic study, we collected real ambient flow data from an office building and simulated a periodic occupancy profile with a peak occupancy of 275 during weekdays and 75 during weekends. Using these two inputs we simulate the dynamics of CO_2 using the model (2.3). We corrupt the CO_2 and ambient flow signals with small amounts of Gaussian noise (standard deviation of 5% of the mean). The signals are presented in Figure 1. The building under study has a total volume of $V = 1.6 \times 10^6$ cf. We assume a CO_2 generation rate per occupant of $G = 0.011$ cfm [14]. The maximum occupancy rate is set to 300 for weekdays and 100 for weekends. We use this to construct piecewise constant bounds L_{noc}, U_{noc} for the occupancy signal. The lower and upper bounds for the input flow are set to 10,000 and 50,000 cfm, respectively. The lower and upper bounds for the CO_2 concentration are set to 400 and 600, respectively. In this first set of experiments we consider a data set with 1,280 time steps and step sizes of 30 minutes (26 days). We run the estimator using the entire horizon. In a later set of experiments we analyze the effect of using a receding horizon approach. The number of expansion coefficients was set to $n_f^i = 140$ for each signal $i \in \{noc, q_{in}\}$.

4.1.1 Regularization Effect

We run the MHE estimator with and without Fourier regularization. The inferred occupancy signals are presented in Figure 2, and the ambient flow signals are presented in Figure 3. As can be observed, slight amounts of noise make the estimates highly volatile. *We emphasize that volatility in estimates is obtained even if the variance of the sensor errors is known.* This indicates that the signals are weakly observable. In other words, slight perturbations on the measurements induce large deviations in

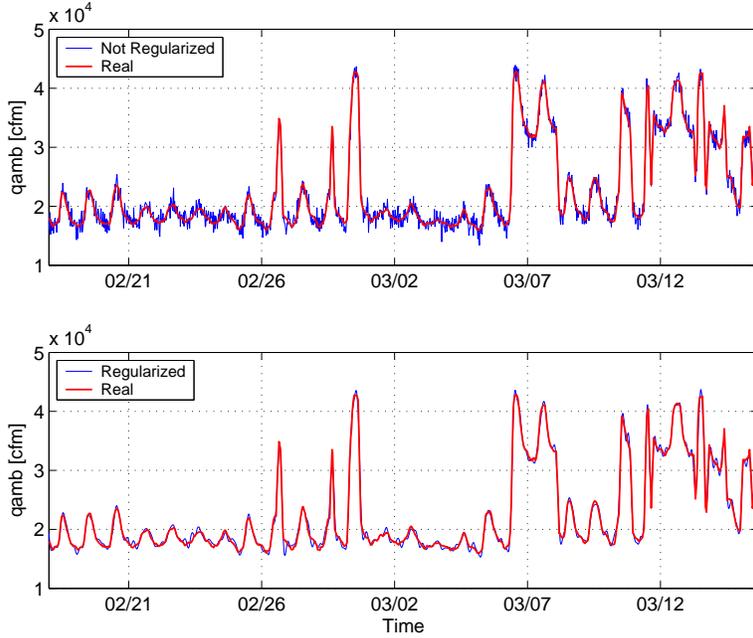


Figure 3: Real and inferred inlet flow signals for MHE under no regularization (top). Real and inferred signals for MHE with regularization (bottom).

the estimates. The reason is the lack of sufficient sensor excitation required infer the entire signal dynamics. This conclusion is reinforced by the fact that the estimates are much more volatile around the flat peaks compared to the transients. The regularization procedure is effective at eliminating this volatility. In addition, we can see that regularized MHE recovers the shape of the occupancy peaks.

In Figure 4 we present cumulative distributions (performance profiles) for the estimation error of the occupancy and flow signals. The horizontal dashed lines indicate the 95% level (i.e., 95% of the time steps). As can be seen, the regularized MHE approach gives errors of less than 10 occupants for 95% of the time. Relative to the peak occupancy of 275 occupants, this represents an error of 3.6%. The unregularized approach has errors of less than 50 occupants 95% of the time and a relative error of 18%. The regularized approach yields an improvement of 80% in the 95% error threshold. For the ambient flow the error obtained with the regularized approach is less than 1,200 cfm for 95% of the time while the error of the unregularized variant is less than 2,000 cfm for 95% of the time. This is an improvement of 40%. This indicates that the benefit is more pronounced for the occupancy signal because this is not measured, as opposed to the flow signal.

4.1.2 Robustness to Noise

To demonstrate the robustness induced by regularization, we performed an experiment with the base level of noise and an experiment with high level of noise. The high noise case contains noise levels of 10% relative to the mean. These levels are twice as large as those of the base noise case. In Figure 5 we present performance profiles for MHE with and without regularization for both cases. For the

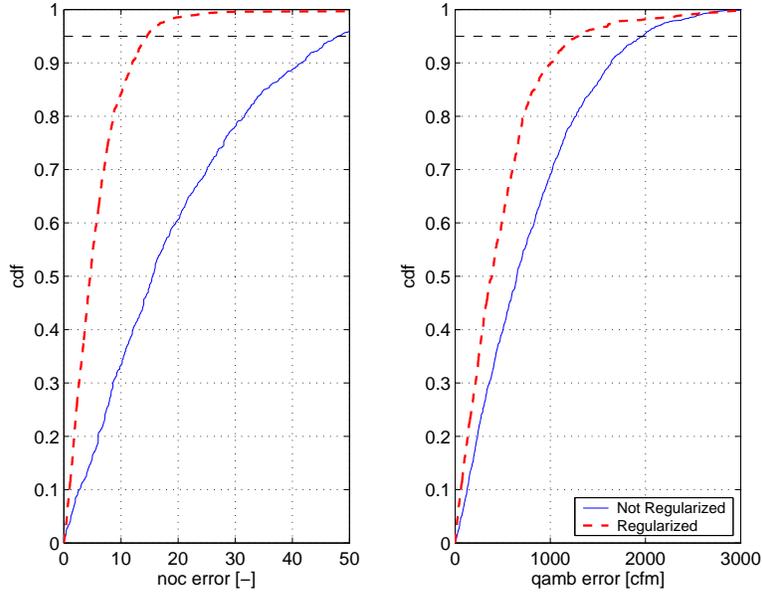


Figure 4: Real and inferred occupancy signals for MHE with and without regularization (left). Real and inferred input flow signals for MHE with and without regularization (right).

unregularized case, we can see that the cumulative probability of an error of less than 50 occupants goes down from 95% to 80%. For the regularized case it remains higher than 99% for both cases. For an error of 20 occupants and below, the cumulative probability of the unregularized estimator goes down from 60 to 45%. For the regularized estimator, the cumulative probability goes down from 98 to 90%. The benefit in robustness is significant.

4.1.3 Number of Expansion Terms

We explore the effect of the number of expansion coefficients on the performance of MHE. We ran a sequence of experiments with a number of expansion coefficients in the range of [10, 250] for the high noise case. In Figure 6 we present the estimator cost as a function of the number of expansion coefficients. As expected, the cost decreases as the number of coefficients is increased. In Figure 7 we compare the performance of MHE with different numbers of expansion coefficients. As can be seen, the best performance (in terms of signal recovery error) is obtained for the case with 150 coefficients. Note also that increasing or decreasing the number of coefficients leads to decreased performance, so an optimum exists.

In a real implementation scenario the true occupancy and flow signals are not known. Consequently, it is not possible to identify a suitable number of expansion coefficients as a function of signal recovery. This situation is common in buildings where obtaining ground truth data for occupancy can be highly expensive. In typical regularization procedures such as Tikhonov regularization, the weight of the regularization term is chosen as the point closest to the origin in the cost-weight curve (in our case the cost-number of coefficients curve). In our case, we have found that this approach

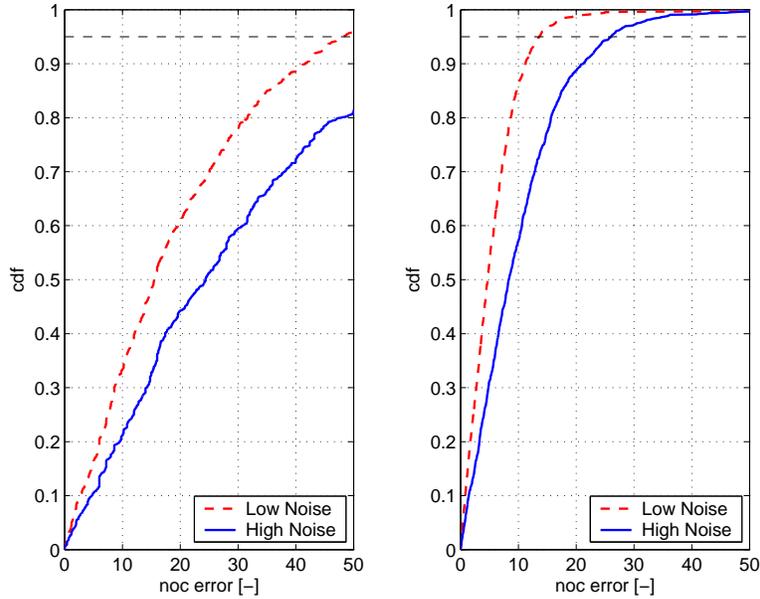


Figure 5: Real and inferred occupancy signals for MHE without regularization (left). Real and inferred occupancy signals for MHE with regularization (right).

gives poor signal recoveries.

To identify a suitable number of expansion coefficients, we constructed a curve of the estimator cost against the variance of the recovered signals. This is presented in Figure 8. This approach is based on the observation that the variance of the recovered signal tends to decrease rapidly as we decrease the number of expansion coefficients. We have observed that the decrease occurs at a faster rate than the corresponding increase in cost. As can be seen in Figure 8 this is indeed the case. The curve indicates that starting at around 140 coefficients, the variance begins decreasing rapidly with decreasing number of coefficients. We have found that the best estimator performance in terms of signal recovery error also occurs at around 140 coefficients, so the approach seems consistent. We emphasize, however, that these observations are empirical in nature and thus further studies on this topic are needed.

4.1.4 Computational Performance and Horizon Length

In Table 1 we present the total CPU times for an estimator with an increasing number of expansion coefficients. These times are for a formulation with a horizon of 1,280 time steps (full problem with horizon of 26 days). The problems were solved by using IPOPT [24] using default options and using MA57 as linear algebra solver. The problem is small in terms of number of variables and constraints. The problem size increases in the number of variables only with the number of coefficients. The number of equality and inequality constraints remains the same. Note, however, that the number of nonzeros per row in the Jacobian matrix increases rapidly, from 69 for the smallest case to 269 to the largest case. These numbers are well beyond typical numbers in other applications, which

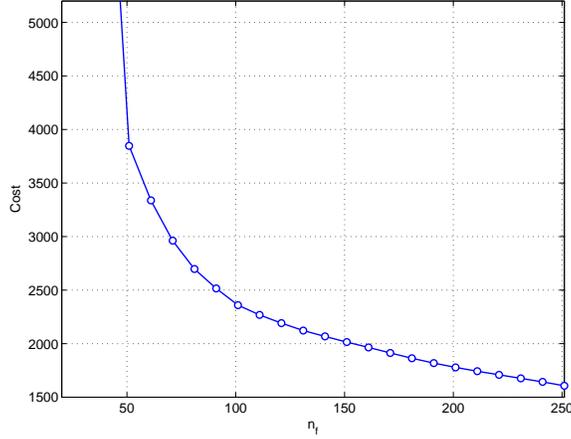


Figure 6: Estimator cost as function of number of expansion coefficients.

Table 1: Dimensions of MHE problem and solution time as a function of number of expansion coefficients.

n_f	n	m_{eq}	m_{in}	nnz	Time [s]
50	5,319	3,837	5,117	2.66×10^5	4.86
100	5,519	3,837	5,117	5.21×10^5	12.08
150	5,719	3,837	5,117	7.77×10^5	28.67
200	5,919	3,837	5,117	10.33×10^5	44.05

range from 5 to 10 nonzeros per row [30]. This is a clear indication that the density of the Jacobian increases significantly with the number of coefficients. This is also reflected by the solution time, which increases nearly cubically with the number of coefficients.

We have observed that performance of MHE does not deteriorate significantly with the horizon length. This is attributed to the inherent periodicity of the signals. To illustrate this, we compare the performance profiles of the MHE estimator with horizons of 7, 6, and 1 days. These correspond to horizons with 336, 288, and 48 time steps of 30 minutes, respectively. The number of coefficients is adjusted to remain proportional with the horizon length, which is key in maintaining estimator robustness and also reducing computational time. For example, the number of coefficients for a 7 day horizon is 34, while that for 1 day is 5.

In Figure 9 we see that the signal recovery performance is not deteriorated significantly over the entire set of 26 days. The 95% error threshold remains below 30 occupants and 2,000 cfm for the case with a 1-day horizon. We have not found appreciable improvements in performance for horizons beyond 7 days. The improvements in computational time, however, are significant. For a 7 day horizon the number of coefficients is 34 and the total time 2.54 seconds per problem compared with the 28.67 seconds required for the full problem using a 26 day horizon.

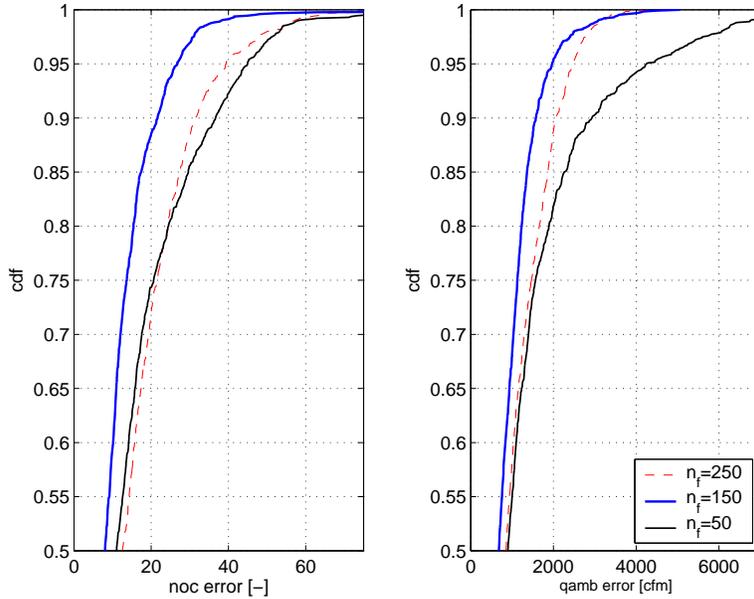


Figure 7: Performance profiles as a function of number of expansion coefficients.

4.2 Real Data

We test the performance of MHE using real sensor readings from the heating, ventilation, and air conditioning (HVAC) system of the Advanced Photon Source office building located at Argonne National Laboratory (see Figure 10). The CO₂ sensor is located in the return duct of the HVAC system. Estimates of the ambient flow are obtained from the sensor readings of the supply duct flows and the damper position of the economizer. We do not have real-time data available on the number of occupants of the building because, as previously mentioned, this is hard to obtain. Operational experience, however, indicates that the peak occupancy fluctuates between 300 and 400 occupants on a typical weekday. The building was in fact designed assuming a peak occupancy of 400.

In this study, we demonstrate that the CO₂ model coupled with regularized MHE gives reasonable estimates (in terms of magnitude and behavior) of the total building-wide occupancy level. The results are presented in Figure 11. In the top two panels we present the CO₂ and ambient flow signals. The solid line is a smoothed signal used to highlight the existing noise in the signals, particularly at peak times. As can be seen, the peak CO₂ levels are decreased as the ambient flow levels are increased. The data is consistent.

The bottom panel presents the estimated occupancy profiles. The MHE estimates remain consistently within the expected range of 300 to 400 and behave in a consistent manner. In particular, note that the peak occupancy levels remain fairly constant despite the wide variability of CO₂ levels and ambient flow. For instance, for the two consecutive weekdays of February 28 and March 1, the occupancy estimates remain at the same level despite the drastic difference in ambient flows. For the consecutive weekdays of March 6 and March 7, the ambient flows and carbon dioxide remain at the same level, and the occupancy profiles are nearly identical, as expected. We also note that MHE

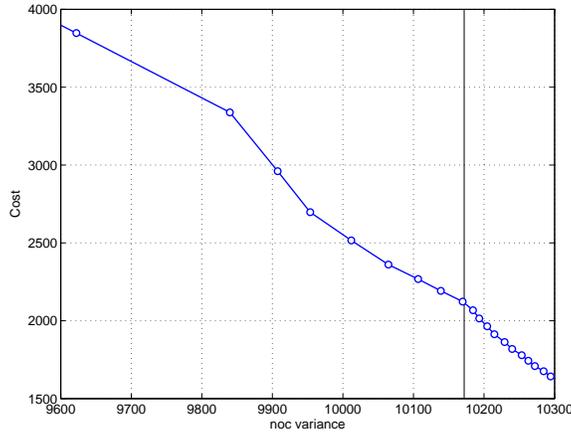


Figure 8: Signal variance against estimator cost. Vertical line indicates case with $n_f = 140$ expansion coefficients.

reconstructs the peak behavior in a consistent manner. In particular, the shape of the peaks indicates an early peak near noon every day followed by a second peak around 4 p.m. This is typical observed behavior in office buildings.

To demonstrate the advantages of regularization we compare the peak occupancy profiles for 5 days for regularized and unregularized MHE. The results shown in the left panel of Figure 12 indicate that the occupancy inferred by unregularized MHE varies strongly and in a manner inconsistent with typical occupancy behavior. The variations are due to the high sensitivity to noise. In particular, in the fourth peak multiple high-frequency peaks and strong occupancy transients are present that are not typical of this building. In particular, in the second peak the estimator indicates a decay of nearly 125 occupants within an hour (30% of the maximum occupancy), which is not typical of this office building. The estimates of regularized MHE are more stable and consistent with typical behavior. These differences in performance are reinforced in the bottom graph where we plot the incremental occupancy per time step (30 min). As can be seen, unregularized MHE leads to increments above 150 occupants in 30 min. In addition, high-frequency variations of occupancy become evident in periods of low occupancy. This volatility is significantly reduced through regularization.

5 Conclusions and Future Work

We have presented a regularization approach for moving horizon estimation motivated by occupancy estimation in building systems. The approach forces input signals to fit a smooth expansion of unknown shape and the expansion coefficients are estimated simultaneously with the states. We demonstrate that the approach leads to significant improvements in robustness in the presence of sensor noise and is particularly useful for recovering weakly observable signals. In addition, it is useful for errors-in-variables formulations that reconstruct measured input signals. As part of future work, we will need to perform additional studies with different building models incorporating thermal and carbon dioxide balances and natural ventilation effects. In addition, a more systematic way

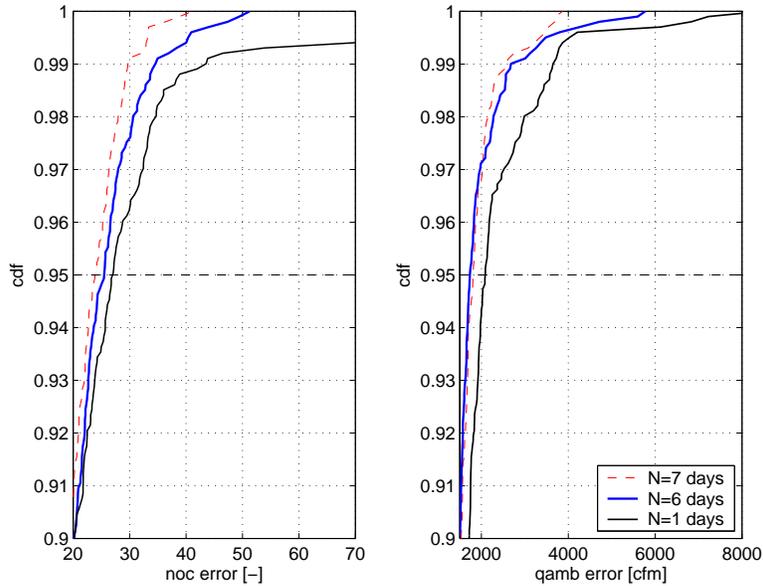


Figure 9: Performance of MHE with Fourier regularization under different estimation horizons.

is needed to determine appropriate number of expansion coefficients.

Acknowledgments

This work was supported by the U.S. Department of Energy, under Contract No. DE-AC02-06CH11357. The author would like to thank Yeonsook Heo for collecting the sensor data used in the real case study.

References

- [1] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44(7):1753–1765, 2008.
- [2] DJ Clements-Croome, HB Awbi, Z. Bako-Biro, N. Kochhar, and M. Williams. Ventilation rates in schools. *Building and Environment*, 43(3):362–367, 2008.
- [3] G. Ferrari-Trecate, D. Mignone, and M. Morari. Moving horizon estimation for hybrid systems. *IEEE Transactions on Automatic Control*, 47(10):1663–1676, 2002.
- [4] E. L. Haseltine and J. B. Rawlings. Critical evaluation of extended Kalman filtering and moving horizon estimation. *Ind. Eng. Chem. Res.*, 44:2541–2460, 2005.



Figure 10: Advanced Photon Source office building at Argonne National Laboratory (reproduced with permission).

- [5] J.D. Hedengren, K.V. Allsford, and J. Ramlal. Moving horizon estimation and control for an industrial gas phase polymerization reactor. In *American Control Conference, 2007. ACC'07*, pages 1353–1358. IEEE, 2007.
- [6] Y. Heo and V.M. Zavala. Gaussian process modeling for measurement and verification of building energy savings. *Energy and Buildings*, 2012.
- [7] J. Hutchins, A. Ihler, and P. Smyth. Modeling count data from multiple sensors: a building occupancy model. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2007.*, pages 241–244. IEEE, 2007.
- [8] S. Ito and H. Nishi. Estimation of the number of people under controlled ventilation using a CO_2 concentration sensor. In *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*, pages 4834–4839. IEEE, 2012.
- [9] T. Kraus, P. Kuhl, L. Wirsching, H.G. Bock, and M. Diehl. A moving horizon state estimation algorithm applied to the Tennessee Eastman benchmark process. In *IEEE International Conference on multisensor Fusion and Integration for Intelligent Systems*, pages 377–382. IEEE, 2006.
- [10] F.V. Lima, M.R. Rajamani, T.A. Soderstrom, and J.B. Rawlings. Covariance and state estimation of weakly observable systems: application to polymerization processes. *IEEE Transactions on Control Systems Technology*, In Press, 2012.
- [11] F.V. Lima and J.B. Rawlings. Nonlinear stochastic modeling to improve state estimation in process monitoring and control. *AIChE Journal*, 57(4):996–1007, 2010.
- [12] T. Lu, X. Lü, and M. Viljanen. A novel and dynamic demand-controlled ventilation strategy for CO_2 control and energy saving in buildings. *Energy and Buildings*, 43(9):2499–2508, 2011.

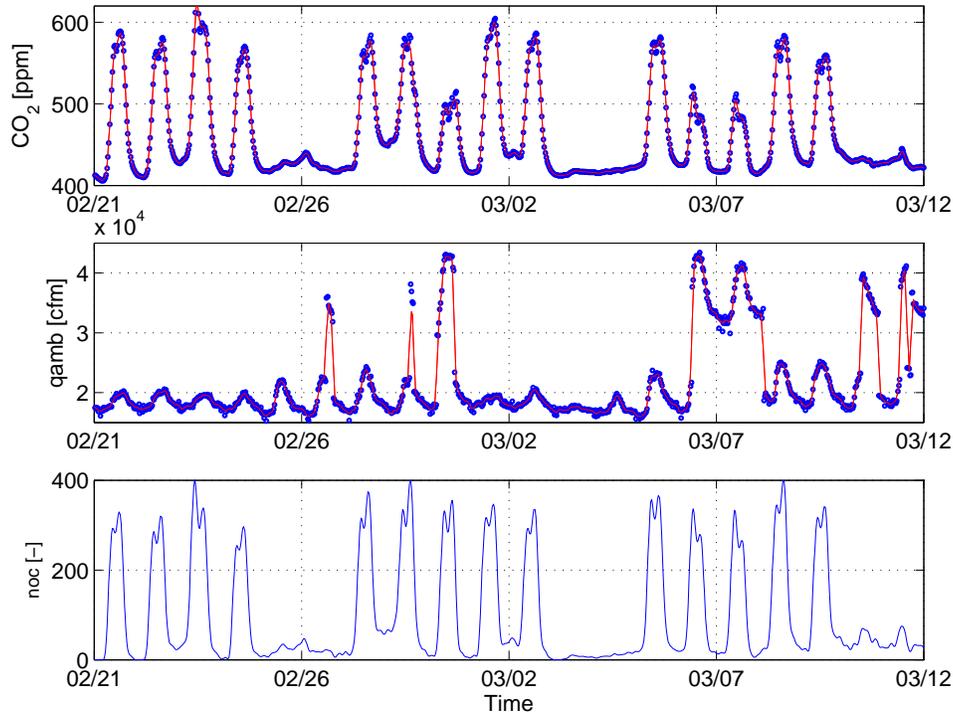


Figure 11: Carbon dioxide and ambient flow sensor signals (top two panels). Estimated occupancy signal using MHE with Fourier regularization (bottom panel).

- [13] S. Meyn, A. Surana, Y. Lin, S.M. Oggianu, S. Narayanan, and T.A. Frewen. A sensor-utility-network method for estimation of occupancy in buildings. In *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 1494–1500. IEEE, 2009.
- [14] S.A. Mumma. Transient occupancy ventilation by monitoring CO_2 . *ASHRAE IAQ Applications*, pages 21–23, 2004.
- [15] B.J. Odelson, M.R. Rajamani, and J.B. Rawlings. A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308, 2006.
- [16] A.A. Poyton, M.S. Varziri, K.B. McAuley, P.J. McLellan, and J.O. Ramsay. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & Chemical Engineering*, 30(4):698–708, 2006.
- [17] M.R. Rajamani and J.B. Rawlings. Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. *Automatica*, 45(1):142–148, 2009.
- [18] J. Ramsay. *Functional data analysis*. Wiley Online Library, 2005.
- [19] C.V. Rao, J.B. Rawlings, and D.Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, 2003.

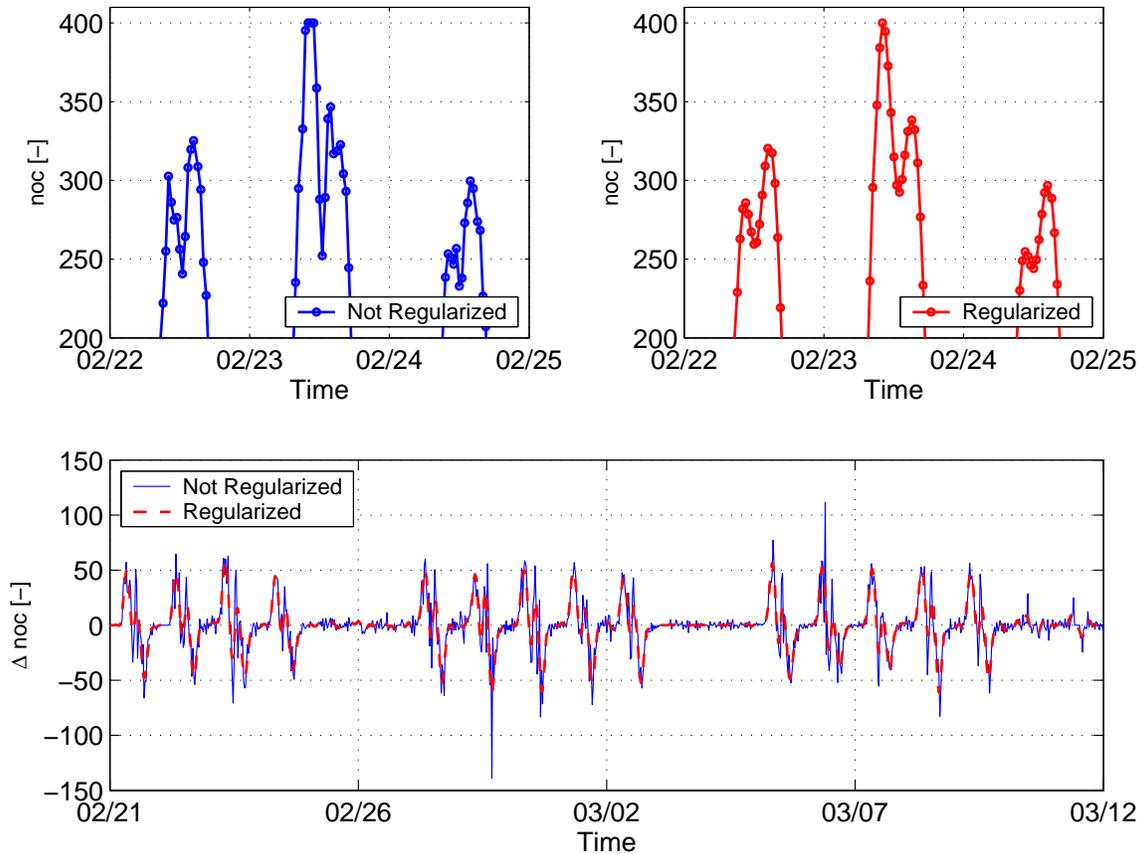


Figure 12: Estimated occupancy signals during peak times for 3 days. Unregularized MHE (top left) and regularized MHE (top right). Occupancy change rate for regularized and unregularized MHE (bottom).

- [20] M. Santamouris, A. Synnefa, M. Assimakopoulos, I. Livada, K. Pavlou, M. Papaglastra, N. Gaitani, D. Kolokotsa, and V. Assimakopoulos. Experimental investigation of the air flow and indoor carbon dioxide concentration in classrooms with intermittent natural ventilation. *Energy and Buildings*, 40(10):1833–1843, 2008.
- [21] J. Široký, F. Oldewurtel, J. Cigler, and S. Prívvara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88(9):3079–3087, 2011.
- [22] B.J. Spivey, J.D. Hedengren, and T.F. Edgar. Constrained nonlinear estimation for industrial process fouling. *Industrial & Engineering Chemistry Research*, 49(17):7824–7831, 2010.
- [23] R. Tomastik, S. Narayanan, A. Banaszuk, and S. Meyn. Model-based real-time estimation of building occupancy during emergency egress. *Pedestrian and Evacuation Dynamics 2008*, pages 215–224, 2010.

- [24] A. Wächter and L. T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106:25–57, 2006.
- [25] H.T. Wang, Q.S. Jia, C. Song, R. Yuan, and X. Guan. Estimation of occupancy level in indoor environment based on heterogeneous information fusion. In *IEEE Conference on Decision and Control.*, pages 5086–5091. IEEE, 2010.
- [26] S. Wang. Dynamic simulation of building vav air-conditioning system and evaluation of emcs on-line control strategies. *Building and Environment*, 34(6):681–705, 1999.
- [27] S. Wang, J. Burnett, and H. Chong. Experimental validation of co2-based occupancy detection for demand-controlled ventilation. *Indoor and Built Environment*, 8(6):377–391, 1999.
- [28] S. Winsberg and J.O. Ramsay. Monotonic transformations to additivity using splines. *Biometrika*, 67(3):669–674, 1980.
- [29] D.P. Word, D.A.T. Cummings, D.S. Burke, S. Iamsirithaworn, and C.D. Laird. A nonlinear programming approach for estimation of transmission parameters in childhood infectious disease using a continuous time model. *Journal of the Royal Society Interface*, 9(73):1983–1997, 2012.
- [30] V. M. Zavala and L. T. Biegler. Large-scale parameter estimation in low-density polyethylene tubular reactors. *Ind. Eng. Chem. Res.*, 45:7867–7881, 2006.
- [31] V.M. Zavala. Stability analysis of an approximate scheme for moving horizon estimation. *Computers & Chemical Engineering*, 34(10):1662–1670, 2010.
- [32] V.M. Zavala and L.T. Biegler. Optimization-based strategies for the operation of low-density polyethylene tubular reactors: Moving horizon estimation. *Computers & Chemical Engineering*, 33(1):379–390, 2009.

The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory (“Argonne”) under Contract No. DE-AC02-06CH11357 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.