Application-oriented input design for room occupancy estimation algorithms

A. Ebadat, D. Varagnolo, G. Bottegal,
B. Wahlberg, *Fellow, IEEE* and K. H. Johansson, *Fellow, IEEE*

***Abstract*— We consider the problem of occupancy estimation in buildings using available environmental information. In particular, we study the problem of how to collect data that is informative enough for occupancy estimation purposes. We thus propose an application-oriented input design approach for designing the ventilation signal to be used while collecting the system identification datasets. The main goal of the method is to guarantee a certain accuracy in the estimated occupancy levels while minimizing the experimental time and effort. To take into account potential limitations on the actuation signals we moreover formulate the problem as a recursive nonlinear and nonconvex optimization problem, solved then using exhaustive search methods. We finally corroborate the theoretical findings with some numerical examples, which results show that computing ventilation signals using experiment design concepts leads to occupancy estimator performing 4 times better in terms of Mean Square Error (MSE).**

***Index Terms*— Occupancy estimation, CO_2 dynamics, application-oriented input design, minimum-time input design**

I. Introduction

Since knowing occupancy levels leads to more efficient Heating, Venting and Air Conditioning (HVAC) control schemes [1]–[4], the problem of occupancy estimation in buildings has been extensively studied in the literature. In general, there exist two categories of occupancy estimation schemes. The first uses dedicated hardware (e.g., [5], [6]), while the second uses statistical estimation algorithms on top of information already available in the Building Management System (BMS) like ventilation signals and measured CO_2 levels (e.g., [7], [8]). The drawbacks of the methods belonging to the first category are the need for installation and maintenance of dedicated hardware plus privacy concerns, while the drawback within the second strategy of methods is an usually higher estimation error. There has been thus an emerging interest in studying if it is possible to improve the estimation algorithms belonging to this second category.

Literature review: as the scheme in Figure 1 exemplifies, estimating occupancy levels from environmental signals is usually performed by inverting the models of the dynamics. Importantly, CO_2 levels are known to be the signals that are most informative for occupancy estimation purposes. The problem is thus to estimate CO_2 models. White-box models

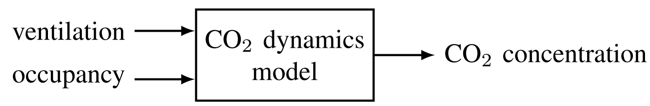

Fig. 1. Example of how to estimate occupancy levels when having a model of the CO_2 dynamics and measurements of the ventilation and CO_2 levels. In this case one may invert the model to find the estimated occupancy as the signal that best explains the measurements through the model.

of the CO_2 dynamics can be obtained directly from physics-based laws like mass-balance equations under well-mixed air conditions, as in [3], [7], [8]. One may also use gray-box models where physics-based concepts are combined with one-to-one correspondences between the model parameter vector and the physical parameters characterizing the room (i.e., room volume and size of the ventilation system), as in [9]. Alternatively one may also use black-box modelling, as in [10]–[16] (with the plethora of citations indirectly reflecting the variegated set of different possibilities for black-box modelling).

Irrespectively of their nature, CO_2 models are always parametrized with unknown parameters that should be estimated through suitable system identification techniques. The quality of the identified model will indirectly reflect into the quality of the consequent occupancy estimates.

An interesting problem is then how to collect model training data that make the consequent occupancy estimates accurate. The problem can be seen as an experiment design [17] problem for the specific occupancy estimation task. Notice that while the generic goal in experiment design is to find an optimal input signal that excites the dynamics of the system, in our setup we focus instead on revealing the systems dynamics that are mostly interesting for the occupancy estimation problem. Therefore, we focus on the problem of application-oriented input design [18], [19].

Statement of contributions: we consider the problem of designing an algorithm that computes how to optimally ventilate a room during the collection of a training dataset to be used for identifying the effects of room occupancy on the CO_2 concentrations in the room. More formally we consider the problem of application-oriented input design when using the nonlinear gray-box model proposed in [9]. The main contributions of this paper are then:

- defining a metric function that measures the degradation

A. Ebadat, B. Wahlberg, and K. H. Johansson are with ACCESS and the Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden. Emails: `{ ebadat | bo | kallej }@kth.se`. D. Varagnolo is with the Division of Signals and Systems, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Sweden. Email: `damiano.varagnolo@ltu.se`. G. Bottegal is with TU Eindhoven, Department of Electrical Engineering, Eindhoven, The Netherlands. Email: `g.bottegal@tue.nl`.

This work was supported by the Swedish Research Council under contract 2016-06079 and the Norrbottens Forskningsråd.

in the accuracy of the estimated occupancy due to the model error. This is instrumental for obtaining a pre-specified accuracy in estimated occupancy based on the defined metric function while minimizing the cost of the identification experiment;

- defining a metric function for the cost of the identification experiment that is proportional to the experiment time. This is because the identification experiment usually involves an accurate knowledge of the occupancy levels, which is usually obtained through temporary people counters. The cost of the experiment is thus proportional to the experiment time;
- introducing a time-recursive algorithm which relies on a short term knowledge of the occupancy signal, so to lessen the computational burden of the optimization problem associated with the experiment design.

The overall strategy is tested in a simulated model which has been identified and validated in [9]. The results show that using the obtained optimal ventilation signal to ventilate the room during the training phase can guarantee the achievement of the required accuracy for the estimated occupancy signals, while satisfying the limitations of the environmental signals. Moreover the occupancy estimators obtained using experiment design concepts are shown to perform at least in silico 4 times better in terms of Mean Square Error (MSE).

Organization of the manuscript: Section II reviews how to perform occupancy estimation using models of the dynamics of the CO_2 concentrations in a room. Section III then presents how to design an identification experiment considering the specific problem of occupancy estimation. Section IV then describes the proposed recursive input design algorithm. Finally, Section V evaluates the effectiveness of the proposed strategy on a numerical example.

II. Occupancy estimation using a gray-box model for the CO_2 concentration levels in a room

Here we review the physics-based CO_2 model that has been proposed and validated in [9], plus explain how this model was used in [20] to construct an occupancy estimator.

A. A physics-based model of CO_2 dynamics

Assuming that the air is well-mixed (see [21]), the CO_2 concentration of a room, denoted by $\overline{c}(t)$, can be modeled from mass-conservation considerations as

$$v\frac{d\overline{c}(t)}{dt} = \left(\dot{Q}^{\text{vent,sup}} + \dot{Q}^{\text{leak,in}}\right)c - \left(\dot{Q}^{\text{vent,exh}} + \dot{Q}^{\text{leak,out}}\right)\overline{c}(t) + g\ o(t). \tag{1}$$

In (1), v is the volume of the room; c is the outdoor air CO_2 concentration, which we assume constant and equal to 420 ppm; $\dot{Q}^{\text{vent,sup}}$ and $\dot{Q}^{\text{vent,exh}}$ are the supply and exhaust mechanical ventilation rates; $\dot{Q}^{\text{leak,in}}$ and $\dot{Q}^{\text{leak,out}}$ are the inflow and outflow air leakages through doors and windows. The product $g\,o(t)$ models the occupants CO_2 generation in the room, where g is the CO_2 generation rate per person and $o(t)$ is the number of occupants at time t.

In the case of balanced ventilation it is reasonable to assume that $\dot{Q}^{\text{vent,sup}} \approx \dot{Q}^{\text{vent}} \approx \dot{Q}^{\text{vent,exh}}$ and that $\dot{Q}^{\text{leak,in}} \approx \dot{Q}^{\text{leak}} \approx \dot{Q}^{\text{leak,out}}$. Under suitable assumptions (see [9]) dynamics (1) can be written as

$$\frac{d\overline{c}(t)}{dt} = \frac{\dot{Q}^{\text{u}}u(t)}{v}\left(c-\overline{c}(t)\right) + \frac{\dot{Q}^{\text{c}}}{v}\left(c-\overline{c}(t)\right) + \frac{g}{v}o(t), \tag{2}$$

with $\dot{Q}^{\text{u}} = \dot{Q}^{\text{vent,max}} - \dot{Q}^{\text{vent,min}}$ and $\dot{Q}^{\text{c}} = \dot{Q}^{\text{vent,min}} + \dot{Q}^{\text{leak}}$ where $\dot{Q}^{\text{vent,max}}$ and $\dot{Q}^{\text{vent,min}}$ are the maximum and minimum airflow through the ventilation system.

Model (2) can then be discretized using backward-Euler rules[1] and sampling time T. Define moreover $c(k) := \overline{c}(k) - c$ and the parameter vector $\boldsymbol{\theta}^T := [\theta_1\ \theta_2\ \theta_3]$, where

$$\theta_1 := \frac{v}{v + T\dot{Q}^{\text{u}}}, \quad \theta_2 := \frac{Tg}{v + T\dot{Q}^{\text{u}}}, \quad \theta_3 := \frac{T\dot{Q}^{\text{c}}}{v + T\dot{Q}^{\text{u}}}.$$

Assuming finally that the measurements of CO_2 are corrupted by additive white Gaussian noise, so that the measured CO_2 concentration $y(k)$ can be expressed as $y(k) = c(k) + e(k)$, the overall CO_2 dynamics can be rewritten as

$$\begin{cases} c(k) = \dfrac{\theta_1}{1+\theta_3 u(k)}c(k-1) + \dfrac{\theta_2}{1+\theta_3 u(k)}o(k) \\ y(k) = c(k) + e(k), \end{cases} \tag{3}$$

where θ_1, θ_2 and θ_3 are model parameters, $u(k)$ is the ventilation signal and $o(k)$ is the occupancy level (see [9] for more details).

Starting from model (3) it is possible to construct a deconvolution-based occupancy estimator as follows: compact the notation by introducing $\boldsymbol{\theta} := [\theta_1, \theta_2, \theta_3]^T$, $\boldsymbol{o} := [o(1), \ldots, o(N)]^T$,

$$a(k\ ;\ \boldsymbol{\theta}) := \frac{\theta_1}{1+\theta_3 u(k)}, \quad \text{and} \quad b(k\ ;\ \boldsymbol{\theta}) := \frac{\theta_2}{1+\theta_3 u(k)},$$

so that (3) becomes

$$c(k) = a(k\ ;\ \boldsymbol{\theta})c(k-1) + b(k\ ;\ \boldsymbol{\theta})o(k). \tag{4}$$

Expanding recursively (4) back in time, and defining the auxiliary variables

$$\widetilde{c}(k) := c(k) - c(0)\prod_{\tau=0}^{k-1} a(k-\tau\ ;\ \boldsymbol{\theta}),$$
$$B(k, k-h\ ;\ \boldsymbol{\theta}) := b(k-h\ ;\ \boldsymbol{\theta})\prod_{\tau=0}^{h-1} a(k-\tau\ ;\ \boldsymbol{\theta})$$

(with the convention that $\prod_{\tau=0}^{-1}\star = 1$ for every $\star$), it is possible to express, for any finite N, the Minimum Variance Unbiased (MVU) estimator of the CO_2 levels $c(k)$ as

$$\widehat{c}\,(k;\boldsymbol{\theta},\boldsymbol{o}) := y(0)\prod_{\tau=0}^{k-1} a(k-\tau\ ;\ \boldsymbol{\theta}) + \left[B(k,1\ ;\ \boldsymbol{\theta}),\ \ldots,\ B(k,k\ ;\ \boldsymbol{\theta})\right]\boldsymbol{o}. \tag{5}$$

From (5) we will then derive an estimator of $\boldsymbol{o}$ in Section II-C.

[1] This choice is motivated by the fact that the backward Euler discretization led to better identification and estimation performance.

B. Prediction Error Method (PEM)-based identification

Estimating $\boldsymbol{o}$ from (5) requires an estimate of the parameter vector $\boldsymbol{\theta}$. Following [22], in this section we describe how to estimate this vector using a PEM approach plus study the statistical properties of this estimator.

Assume then the availability of measurements of the occupancy signal thanks to a temporary people counter inside the room, so that the dataset includes the occupancy levels $o(k)$, the measured CO_2 concentrations $y(k)$, and the ventilation actuation levels $u(k)$ available, i.e., $\mathcal{D} := \{\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{o}\}$. Let $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ be a generic guess of the model parameters in (3). $\boldsymbol{\theta}$ and $\mathcal{D}$ jointly define the difference between the recorded output $\boldsymbol{y}$ in $\mathcal{D}$ and the predicted output $\widehat{c}$ computed through model (5). Defining the prediction error as $\varepsilon(k, \boldsymbol{\theta}) := y(k) - \widehat{c}(k; \boldsymbol{\theta}, \boldsymbol{o})$ and following the PEM paradigm, we measure the quality of an estimated $\boldsymbol{\theta}$ based on the predictions $\varepsilon(k, \boldsymbol{\theta})$. Given our Gaussian assumptions on $e(k)$ in (3) we thus exploit quadratic costs, i.e.,

$$\widehat{\boldsymbol{\theta}}_N = \arg\min_{\theta \in \Theta} V_N(\boldsymbol{\theta}, \mathcal{D}), \quad V_N(\boldsymbol{\theta}, \mathcal{D}) := \frac{1}{2N} \sum_{k=1}^{N} \varepsilon^2(k, \boldsymbol{\theta}),$$

with $\Theta \subset \mathbb{R}^{n_\theta}$ defined such that for all $\theta \in \Theta$ the model (3) exists and is stable.

An important asymptotic property of a PEM estimator (in the sample size N) is that, assuming that there exists an ideal parameter vector $\boldsymbol{\theta}_0$ from which the data is generated, and under additional mild assumptions,

$$\left(\mathcal{I}_F^N(\boldsymbol{\theta}_0)\right)^{\frac{1}{2}} \left(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I) \tag{6}$$

where d means in distribution and where $\mathcal{I}_F^N$ is the Fisher Information Matrix (FIM), i.e.,

$$\mathcal{I}_F^N(\boldsymbol{\theta}_0) := \frac{1}{\sigma_e^2} \sum_{k=1}^{N} \mathbb{E}_e \left\{ \psi(k, \boldsymbol{\theta}_0) \, \psi(k, \boldsymbol{\theta}_0)^T \right\} \tag{7}$$

(see [23] for more details on the assumptions and derivations). We assume $\mathcal{I}_F^N$ invertible. The expectation above is taken with respect to the measurement noise e, and

$$\begin{aligned}\psi(k, \boldsymbol{\theta}_0) \quad &:= \left[\frac{d\widehat{c}(k; \boldsymbol{\theta}, \boldsymbol{o})}{d\theta_1}, \ \frac{d\widehat{c}(k; \boldsymbol{\theta}, \boldsymbol{o})}{d\theta_2}, \ \frac{d\widehat{c}(k; \boldsymbol{\theta}, \boldsymbol{o})}{d\theta_3} \right]^T \\ &= \left[-\frac{d\varepsilon(k; \boldsymbol{\theta})}{d\theta_1}, \ -\frac{d\varepsilon(k; \boldsymbol{\theta})}{d\theta_2}, \ -\frac{d\varepsilon(k; \boldsymbol{\theta})}{d\theta_3} \right]^T\end{aligned}$$

(see [17], [22] for details on this derivation). Leveraging on (6), we can construct a confidence set around the true parameters $\boldsymbol{\theta}_0$ for which, as $N \to \infty$,

$$\mathcal{E}_{\text{id}}(\boldsymbol{\theta}_0, \alpha) := \left\{ \boldsymbol{\theta} \ : \ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mathcal{I}_F^N(\boldsymbol{\theta}_0) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \le \chi_\alpha^2(n_{\boldsymbol{\theta}}) \right\} \tag{8}$$

where $\chi_\alpha^2(n_{\boldsymbol{\theta}})$ is the α-percentile of the χ^2-distribution with $n_{\boldsymbol{\theta}}$ degrees of freedom. I.e., for large N $\widehat{\boldsymbol{\theta}}_N \in \mathcal{E}_{\text{id}}(\boldsymbol{\theta}_0, \alpha)$ with probability at least approximately α. We then call $\mathcal{E}_{\text{id}}(\boldsymbol{\theta}_0, \alpha)$ the *identification set*.

For a given dataset $\mathcal{D}$ we can estimate $\mathcal{I}_F^N(\boldsymbol{\theta}_0)$ as

$$\mathcal{I}_F^N(\boldsymbol{\theta}_0) \approx I_F^N(\boldsymbol{\theta}_0) := \frac{1}{\sigma_e^2} \sum_{k=1}^{N} \psi(k, \boldsymbol{\theta}_0) \, \psi(k, \boldsymbol{\theta}_0)^T. \tag{9}$$

Notice that all the approximations above are based on the fact that N is "sufficiently large". Translating this into practical rules to check the validity of these approximations is well beyond the scope of this paper. We will thus assume that in our framework N will always be large enough.

C. Estimating $\boldsymbol{o}$ using $\widehat{\boldsymbol{\theta}}_N$

Having an estimate $\widehat{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}$, measurements of $y(k)$ (i.e., estimates of $\widehat{c}(k)$) and $u(k)$ for a certain time period $k = 1, \ldots, N$, one can then estimate the occupancy levels $\boldsymbol{o}$ by minimizing the error between the CO_2 predicted through model (5) and the measured one. Following [20] the estimator can then be cast as the regularized program

$$\widehat{\boldsymbol{o}}(\widehat{\boldsymbol{\theta}}_N) = \arg\min_{\boldsymbol{o}} \sum_{k=1}^{N} \left(y(k) - \widehat{c}\left(k; \widehat{\boldsymbol{\theta}}_N, \boldsymbol{o}\right) \right)^2 + \lambda \|\Delta \boldsymbol{o}\|_1 \tag{10}$$

where $\Delta \boldsymbol{o}$ is the discrete derivative of the vector $\boldsymbol{o}$ and $\|\cdot\|_1$ is ℓ_1 norm (see [20] for more details).

III. Designing experiments to obtain improved occupancy estimations

Assume $\widehat{\boldsymbol{o}}\left(\widehat{\boldsymbol{\theta}}_N\right)$ to be the occupancy estimate returned by (10) and where we explicit that the solution is a function of the identified parameter $\widehat{\boldsymbol{\theta}}_N$. Let our application-oriented metric function be

$$V_{\text{app}}\left(\widehat{\boldsymbol{\theta}}_N, \boldsymbol{\theta}_0\right) := \mathbb{E}_{\boldsymbol{u}} \left\{ \frac{1}{N} \left\| \widehat{\boldsymbol{o}}\left(\widehat{\boldsymbol{\theta}}_N\right) - \widehat{\boldsymbol{o}}(\boldsymbol{\theta}_0) \right\|_2^2 \right\}, \tag{11}$$

where we implicitly assume that estimator $\widehat{\boldsymbol{o}}(\cdot)$ performs at its best whenever $\widehat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}_0$ and where the expected value is with respect to the ventilation signals used while estimating the occupancy. The metric function (11) is called *application function* and measures how much we pay in terms of occupancy estimation accuracy if we do not identify the true parameters. In other words, the accuracy of $\widehat{\boldsymbol{\theta}}_N$ reflects on the amplitude of $V_{\text{app}}\left(\widehat{\boldsymbol{\theta}}_N, \boldsymbol{\theta}_0\right)$.

Our focus is now to find how to estimate $\widehat{\boldsymbol{\theta}}_N$ (i.e., how to collect the training dataset $\mathcal{D}$) so that a pre-specified accuracy for the estimated occupancy levels can be guaranteed. Notice that we formulate the problem as an *application-oriented input design* problem (see [24]). More explicitly, considering that $\mathcal{D} := \{\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{o}\}$, we focus on designing the optimal ventilation signal $\boldsymbol{u}$ considering the occupancy $\boldsymbol{o}$ as a non-controllable (and thus given) input and the CO_2 levels $\boldsymbol{y}$ the output of the system.

To specify what "optimal ventilation" means we then use the metric function V_{app} to define the set of admissible parameters $\boldsymbol{\theta}$ as follows: assume that a pre-defined accuracy γ^{-1} has been chosen by the user. Then our requirement is

$$\widehat{\boldsymbol{\theta}}_N \in \Theta_{\text{app}}(\boldsymbol{\theta}_0, \gamma) := \left\{ \boldsymbol{\theta} \ : \ V_{\text{app}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le \gamma^{-1} \right\}, \tag{12}$$

where $\Theta_{\text{app}}(\boldsymbol{\theta}_0, \gamma)$ is the so-called *application set*.

The problem is thus how to design $\boldsymbol{u}$ so that the dataset $\mathcal{D}$ guarantees that the PEM estimate $\widehat{\boldsymbol{\theta}}_N$ is s.t. $\widehat{\boldsymbol{\theta}}_N \in \Theta_{\text{app}}(\boldsymbol{\theta}_0, \gamma)$.

A. Knowledge of true system parameters

One may set up the experiment design problem by laddering on the FIM (7), the identification set (8) and the application set (12). These quantities, though, depend on the true parameters $\boldsymbol{\theta}_0$, that $\boldsymbol{\theta}_0$ is not known a-priori. To circumvent this problem there are then mainly two approaches: either performing robust experiment design (e.g., [25], [26]) or adaptive input design (e.g., [27]).

In this paper we use an initial estimate of the true parameter vectors which can be obtained through either a cheap identification experiment or physical knowledge of the system. We denote this initial estimation by $\widehat{\boldsymbol{\theta}}_0$. We replace $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}}_0$ in all the necessary expressions in the rest of this paper. We assume one can make the designed input robust to the uncertainties in the initial estimates using the robust application-oriented input design approach in [26, Ch. 9].

B. Application-oriented input design

We define our application-oriented input design problem in general terms as:

$$\begin{array}{ll} \underset{\boldsymbol{u}}{\text{minimize}} & \text{Experimental Cost} \\ \text{subject to} & \widehat{\boldsymbol{\theta}}_N \in \Theta_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \gamma) \\ & \boldsymbol{u} \in \text{Input Constraints} \\ & \boldsymbol{y} \in \text{Output Constraints.} \end{array} \tag{13}$$

In the following subsections we then explicit the quantities mentioned in (13) in mathematical terms.

1) Experimental cost: collecting the dataset $\mathcal{D}$ to perform system identification requires temporary installation and maintenance of people counting devices, and this makes it advantageous to make the system identification phase as short as possible. To this end, we consider the length of the identification experiment N as the experiment cost. To put it another way, we aim at fining the minimum time required to fulfill the constraints.

2) Model quality constraints: the model quality constraint requires

$$\widehat{\boldsymbol{\theta}}_N \in \Theta_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \gamma). \tag{14}$$

Since $\widehat{\boldsymbol{\theta}}_N$ is a stochastic variable, it is not possible to enforce deterministic bounds on it. We thus replace the constraint in (14) with the identification set $\mathcal{E}_{\text{id}}(\widehat{\boldsymbol{\theta}}_0, \alpha)$ in (8), i.e., replace (14) with

$$\mathcal{E}_{\text{id}}(\widehat{\boldsymbol{\theta}}_0, \alpha) \subseteq \Theta_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \gamma), \tag{15}$$

which ensures $\widehat{\boldsymbol{\theta}}_N \in \Theta_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \gamma)$ with probability at least α assuming N large enough (see also [28] for other formulations of this type of problems).

Unfortunately, though, the set constraint (15) is not necessarily convex: indeed the identification set $\mathcal{E}_{\text{id}}(\widehat{\boldsymbol{\theta}}_0, \alpha)$ is an ellipsoid, as stated in Section II-B, but the application set $\Theta_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \gamma)$ in (12) can be of any shape. The known methods to find a convex approximation of this constraint are the scenario approach [29], [30] and the ellipsoidal approach [31]. In the scenario approach the application set is described by a finite number of samples for which the constraint (14) should be fulfilled. In order to have a good approximation of the set constraint, the number of scenarios must be large enough. Thus, the scenario approach requires several evaluations of the application function.

In this paper we use the ellipsoidal approximation method which employs a second order Taylor expansion of $V_{\text{app}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_0)$ around $\widehat{\boldsymbol{\theta}}_0$. Noting that $V_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0) = 0$ and assuming that $\widehat{\boldsymbol{\theta}}_0 \in \Theta$ and that $V_{\text{app}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_0)$ is twice differentiable, we can indeed write

$$\begin{aligned} V_{\text{app}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_0) &\approx V_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0) + V'_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_0) \\ &\quad + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_0)^T V''_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_0) \\ &= \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_0)^T V''_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_0). \end{aligned}$$

The application set can thus be approximated by

$$\Theta_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \gamma) \approx \left\{ \theta : [\theta - \widehat{\boldsymbol{\theta}}_0]^T V''_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0)[\theta - \widehat{\boldsymbol{\theta}}_0] \le \frac{2}{\gamma} \right\}. \tag{16}$$

Using (16) to define $\Theta_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \gamma)$ instead of (12) implies that the set constraint (15) can be rewritten, after simple manipulations, as

$$\frac{1}{\chi^2_\alpha(n_{\boldsymbol{\theta}})} \mathcal{I}_F^N(\widehat{\boldsymbol{\theta}}_0) \ge \frac{\gamma}{2} V''_{\text{app}}(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0). \tag{17}$$

with $\mathcal{I}_F^N(\widehat{\boldsymbol{\theta}}_0)$ the FIM defined in (7) evaluated at the best available estimation of the true parameters $\widehat{\boldsymbol{\theta}}_0$. We further simplify (17) by replacing $\mathcal{I}_F^N(\widehat{\boldsymbol{\theta}}_0)$ with its approximation $I_F^N(\widehat{\boldsymbol{\theta}}_0)$ defined in (9), which makes it possible to compute the information matrix for a given dataset. In this way the model quality constraint then becomes

$$\frac{1}{\chi^2_\alpha(n_{\boldsymbol{\theta}})} I_F^N\left(\widehat{\boldsymbol{\theta}}_0\right) \ge \frac{\gamma}{2} V''_{\text{app}}\left(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0\right). \tag{18}$$

This approximation can then be transformed into an alternative (and more prone to numerical implementations) version by defining

$$\widetilde{V} := \frac{\gamma \chi^2_\alpha(n_{\boldsymbol{\theta}})}{2} V''_{\text{app}}\left(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0\right).$$

Because $\widetilde{V}$ is at least semi-definite positive, we can indeed write $\widetilde{V} = \widetilde{V}^{1/2}\widetilde{V}^{1/2}$. The requirement (18) thus becomes

$$\widetilde{V}^{-1/2} I_F^N\left(\widehat{\boldsymbol{\theta}}_0\right) \widetilde{V}^{-1/2} \ge I$$

or, equivalently,

$$\lambda_{\min}\left(\widetilde{V}^{-1/2} I_F^N\left(\widehat{\boldsymbol{\theta}}_0\right) \widetilde{V}^{-1/2}\right) \ge 1, \tag{19}$$

where $\lambda_{\min}(X)$ indicates the minimum eigenvalue of a s.d.p. matrix X.

3) Input constraints: the set of constraints on the ventilation signal $u(k)$ is typically composed by:

- time-invariant box-constraints of the form

$$u(k) \in \mathcal{U}, \text{ for all } k,$$

  where $\mathcal{U}$ is the set of admissible values for the ventilation signal. The signal $u(k)$ has the physical meaning of

opening percentage of a ventilation valve. We assume that the opening is a discrete signal and thus the set $\mathcal{U}$ is a discrete set with finite cardinality;
- restrictions on the number of changes in the ventilation signal due to concerns on the physical integrity of the actuators. Assuming the number of possible changes in the ventilation signal is a constant n_c, this constraint can be formulated as

$$\|\Delta \boldsymbol{u}\|_0 \leq n_c,$$

where

$$\begin{aligned}
\Delta \boldsymbol{u} &:= \begin{bmatrix} \Delta u(1) & \dots & \Delta u(N) \end{bmatrix}, \\
\Delta u(k) &:= u(k) - u(k-1),
\end{aligned}$$

and $\|\cdot\|_0$ is the zero norm, i.e., the number of non-zero elements in the vector $\Delta \boldsymbol{u}$;
- considering that the final goal of occupancy estimation is to optimize the performance of controllers and thus save energy consumption, it is also desirable to use a low-energy ventilation signal during the identification experiment. Unfortunately, this is in contradiction with the excitation requirements, since more excitation usually leads to better estimates. We thus capture this intrinsic trade-off with the constraint

$$\|\boldsymbol{u}\|_2^2 \leq (1+\beta)\|\boldsymbol{u}^*\|_2^2$$

where $\boldsymbol{u}^*$ is the minimum-energy input sequence that fulfills the input and output constraints, i.e.,

$$\begin{aligned}
\boldsymbol{u}^* = \quad & \arg\min && \|\boldsymbol{u}\|_2^2 \\
& \text{subject to} && \boldsymbol{u} \in \text{Input Constraints} \\
& && \boldsymbol{y} \in \text{Output Constraints}
\end{aligned} \tag{20}$$

and $\beta \geq 0$ is a scalar that determines how much more energy the experiment-design algorithm is allowed to use for its identification purposes. Notice that if there are no restrictions on the energy usage then one can simply remove this additional constraint from the problem formulation by putting $\beta = \infty$. This formulation is motivated by the idea behind dual control, see [32].

4) Output constraints: restrictions on the output signal $y(k)$ are in general due only to requirements on the quality of the indoor air. However, given our Output Error (OE) with Gaussian noise model (3), it is not possible to impose deterministic constraints on $y(k)$. We thus consider

$$\mathbb{P}\{y(k) \leq y_{\text{max}}\} \geq p_y \quad \text{for every } k,$$

and explain how to deal with this probabilistic constraint in practice in Section IV-B.

C. Reformulation and comments

Given the results above we can rewrite our application-oriented input design strategy in (12) as

$$\begin{aligned}
\underset{\boldsymbol{u},N}{\text{minimize}} \quad & N \\
\text{subject to} \quad & \lambda_{\min}\left(\widetilde{V}^{-1/2} I_F^N\left(\widehat{\boldsymbol{\theta}}_0\right)\widetilde{V}^{-1/2}\right) \geq 1 \\
& \boldsymbol{u} \in \mathcal{U} \\
& \|\Delta \boldsymbol{u}\|_0 \leq n_c \\
& \|\boldsymbol{u}\|_2^2 \leq (1+\beta)\|\boldsymbol{u}^*\|_2^2 \\
& \mathbb{P}\{\boldsymbol{y} \leq y_{\text{max}}\} \geq p_y .
\end{aligned} \tag{21}$$

Importantly, formulating problem (21) requires knowing the estimate of the parameters $\widehat{\boldsymbol{\theta}}_0$, that should thus have been obtained before the design step; the approximated information matrix I_F^N, defined in (9); and the Hessian of the metric function V_{app}, defined in (11).

In their turn computing I_F^N and V''_{app} requires knowing both the occupancy signal $\boldsymbol{o}$ and the CO_2 levels $\boldsymbol{y}$ that will happen during the experiment under design, a clearly infeasible request.

Solving problem (21) poses several challenges:

C1) it is required to know quantities that are not necessary available when the problem should be solved;

C2) the problem is highly non-linear and non-convex, with associated numerical challenges specially for large N.

In the following section we will describe how to tackle these issues through a suitable reformulation of the application-oriented input design problem (21).

IV. Recursive application-oriented input design for occupancy estimation

A. Step 1: address challenge C1

Recall that computing $I_F^N\left(\widehat{\boldsymbol{\theta}}_0\right)$ and $V''_{\text{app}}\left(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_0\right)$ would require knowing both the occupancy $\boldsymbol{o}$ and the CO_2 levels $\boldsymbol{y}$ *that will happen during the datasets collection step*. Even if this sounds impractical, a workaround may be to assume $\boldsymbol{o}$ to be known a-priori thanks to suitable building schedule information (e.g., a-priori knowledge that a specific office will be occupied by a specific number of people at specific times). Knowing then $\boldsymbol{o}$ and an estimated $\widehat{\boldsymbol{\theta}}_0$ one may simulate $\boldsymbol{y}$ as a function of $\boldsymbol{u}$ through (3) and thus get all the necessary information for computing I_F^N and V''_{app}.

This workflow is nonetheless very prone to errors for two reasons: first, the occupancy schedules are intrinsically uncertain; second, the accuracy of the $\boldsymbol{y}$ obtained by simulations through $\widehat{\boldsymbol{\theta}}_0$ and (3) is greatly affected by the accuracy of $\widehat{\boldsymbol{\theta}}_0$ itself. To obtain a more robust experiment design strategy one can then use a recursive input design algorithm based on the following key concepts:

1) assume that at each time t we have knowledge of the past values of the signals, so that it is possible to compute the information that has been obtained from this measured data, i.e., to compute the "approximate FIM up to t"

$$I_F^t(\boldsymbol{\theta}_0) = \frac{1}{\sigma_e^2}\sum_{k=1}^{t} \psi\left(k, \boldsymbol{\theta}_0\right)\psi\left(k, \boldsymbol{\theta}_0\right)^T ; \tag{22}$$

2) choose a time horizon with a length N_u that is sufficiently small for making persistency-based occupancy

forecasts accurate. I.e., make N_u small enough that when given the number of occupiers at time $k = t$ the occupancy levels $o(t+1), \ldots, o(t+N_u)$ are very likely to remain equal to $o(t)$;

3) run an opportunely modified version of (21) that has a time horizon of N_u steps and that designs the future $u(t+1), \ldots, u(t+N_u)$ so to obtain as much information as possible during the period $t+1, \ldots, t+N_u$ (see the program (23) at the end of this algorithm);
4) use the computed ventilation signal $u(t+1), \ldots, u(t+N_u)$ to run the experiment for N_u more samples;
5) at time instant $t + N_u$ check the amount of information that has been collected is enough, i.e., if the experiment design constraint (19) is satisfied (notice that this constraint refers to the whole time horizon of the experiment, and not only to the period $t+1, \ldots, t+N_u$);
6) in case (19) is satisfied stop the experiment, otherwise continue running it.

The optimization problem that we propose to solve recursively for every t in step 3 is thus

$$
\begin{array}{ll}
\underset{\mu, \boldsymbol{u}_t}{\text{maximize}} & \mu \\
\text{subject to} & \lambda_{\min}\left(\left(\widetilde{V}^{-1/2}\left(I_F^{t+N_u}\right)\left(\widehat{\boldsymbol{\theta}}_0\right)\widetilde{V}^{-1/2}\right)\right) \geq \mu \\
& \boldsymbol{u}_t \in \mathcal{U} \\
& \|\Delta \boldsymbol{u}_t\|_0 \leq n_{ct} \\
& \|\boldsymbol{u}_t\|_2^2 \leq (1+\beta)\|\boldsymbol{u}_t^*\|_2^2 \\
& \mathbb{P}\{\boldsymbol{y} \leq y_{\max}\} \geq p_y \ ,
\end{array}
\tag{23}
$$

where

- the variable $\boldsymbol{u}_t = \{u(t+1), \ldots, u(t+N_u)\}$ is the vector of ventilation signal on the considered horizon;
- μ measures the experiment quality, and maximizing μ is equivalent to maximizing the obtained information;
- n_{ct} is the number of allowed changes during the time horizon $[t, t+N_u]$;
- the vector $\boldsymbol{u}_t^* = \{u^*(t+1), \ldots, u^*(t+N_u)\}$ is obtained by solving (20) for the time horizon $[t, t+N_u]$;
- the matrix $I_F^{t+N_u}(\widehat{\boldsymbol{\theta}}_0)$ is given by (9), i.e.,

$$
\begin{aligned}
I_F^{t+N_u}(\widehat{\boldsymbol{\theta}}_0) = & \frac{1}{\sigma_e^2}\sum_{k=1}^{t}\psi\left(k,\widehat{\boldsymbol{\theta}}_0\right)\psi\left(k,\widehat{\boldsymbol{\theta}}_0\right)^T \\
& + \frac{1}{\sigma_e^2}\sum_{k=t+1}^{t+N_u}\psi\left(k,\widehat{\boldsymbol{\theta}}_0\right)\psi\left(k,\widehat{\boldsymbol{\theta}}_0\right)^T .
\end{aligned}
\tag{24}
$$

The first term in the Right Hand Side (RHS) of (24) is computable given the information available up to time t (and can thus be computed recursively), while the second term depends on the decision variable $\boldsymbol{u}_t$.

One main advantage of using the proposed recursive approach is that one can update the previous information and thus compensate for possible uncertainties. The algorithm is eventually summarized in Algorithm 1.

B. Step 2: address challenge C2

To design the optimal ventilation signal in Algorithm 1 one needs to solve (20) and (23) at each iteration. However,

Algorithm 1 Minimum-time application-oriented input design algorithm

1: **Require:** $V''_{\text{app}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$, α, β, γ, $y_{\max}$, p_y, n_{ct}
2: **Initialization:** put $t = 0$, $\mu = 0$ and $I_F^t(\widehat{\boldsymbol{\theta}}_0) = \mathbf{0}$
3: **while** $\lambda_{\min}\left(\left(\widetilde{V}^{-1/2}\left(I_F^{t+N_u}\right)\left(\widehat{\boldsymbol{\theta}}_0\right)\widetilde{V}^{-1/2}\right)\right) \leq 1$ **do**
4: find $\boldsymbol{u}_t^*$ by solving (20)
5: find $\boldsymbol{u}_t$ by solving (23)
6: run the experiment using $\boldsymbol{u}_t$
7: compute $I_F^{t+N_u}$ using measurements and (22)
8: compute $\lambda_{\min}\left(\left(\widetilde{V}^{-1/2}\left(I_F^{t+N_u}\right)\left(\widehat{\boldsymbol{\theta}}_0\right)\widetilde{V}^{-1/2}\right)\right)$

both (20) and (23) are non-linear and nonconvex. Since $\mathcal{U}$ has finite cardinality and the ventilation signal is only allowed to change n_c times, it is possible to reduce the problem to the problem of finding optimal change times and levels instead. Moreover, since the optimization horizon at each iteration has length N_u, one can solve the problem effectively by finding all possible ventilation signals and thus using an exhaustive search approach.

Another challenge in solving the above mentioned optimization problems is evaluating the probabilistic constraint, $\mathbb{P}\{y(k) \leq y_{\max}\} \geq p_y$. We then notice that using the exhaustive search approach, it is possible to evaluate for each possible signal $\{u(k)\}_{k=t}^{t+N_u}$ the probabilistic constraint using the approximation

$$
\mathbb{P}\{y(k) \leq y_{\max}\} = \frac{\sum_{k=1}^{t+N_u} \mathbb{1}\left(y_{\max} - y(k)\right)}{t+N_u},
$$

where $\mathbb{1}(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0 \end{cases}$.

V. Numerical Results

We evaluate our proposed input design approach through a dedicated numerical example where the synthetic data is generated as follows: as for the model parameters, we use the values that have been identified in [9] starting from a dataset collected in a laboratory in KTH, Stockholm, i.e., $\theta_1 = 0.8639$, $\theta_2 = 10.779$ and $\theta_3 = 0.2189$. As for the additive noise in (3) we choose $\sigma_e^2 = 10$, a value that mimics what has been identified in [9] and we consider $T_s = 5$ minutes as the sampling time. The used occupancy pattern is shown in Figure 2.

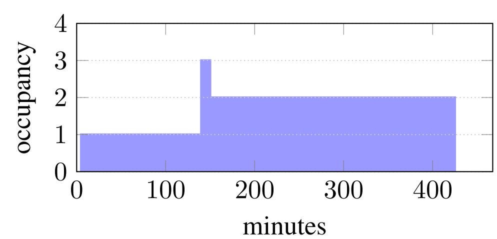

Fig. 2. The occupancy pattern during the input design process.

A. Input design for parameter identification

We then use the proposed approach in Section IV to design the optimal ventilation signal assuming that:

- $N_u = 12$ (1 hour), plus we are allowed to change the ventilation signal once each hour, i.e, $n_{ct} = 1$;
- the estimated parameters should lie in the application set with the probability of at least $\alpha = 98\%$;
- the chosen accuracy is $\gamma = 100$;
- the occupancy pattern is assumed to be known during the time horizons of length N_u;
- the measured output $y(k)$ should not violate $y_{\text{max}} = 120$ ppm with the probability of at least $p_y = 75\%$;
- the set of possible values for the ventilation signal is given by $\mathcal{U} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

We then solve at each iteration the optimization problem (23) through a combinatorial optimization tool. Notice that since we are only allowed to change the ventilation signal once at each iteration, and that the ventilation can only take a finite number of values, we use an exhaustive search.

We then consider two different cases:

Case 1: the variable β is chosen to be infinity which means we assume no upper bound on the energy of the signal being used. In this case the optimal computed ventilation signal is bang-bang, as for example shown in Figure 3. This reflects the intuition that, since there is no constraint on the energy of the ventilation, the optimal strategy is to maximize the excitation by using an on-off strategy. Notice that in this specific case the required excitation is obtained in $t = 425$ minutes, and that the energy of the computed signal is $\|\boldsymbol{u}\|_2^2 = 47.24$.

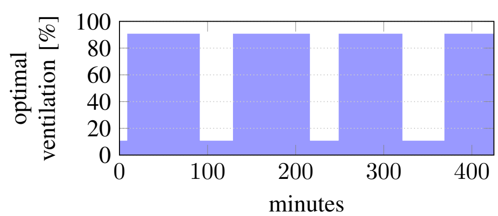

Fig. 3. The obtained optimal ventilation signal for the case $\beta = \infty$.

Case 2: the variable β is chosen to be 0.2, that means that we allow the input-design algorithm to use at most 20% more ventilation energy than the one that would be used if we were not requiring excitation. Figure 4 then shows that in this case the required time to get enough excitation is increased to $t = 605$ minutes. Also this follows the intuition, since the amount of available excitation at each iteration is now limited compared to the previous case. Nonetheless the total energy of the obtained ventilation signal is $\|\boldsymbol{u}\|_2^2 = 47.28$, very close to the energy of the optimal signal in case 1. Somehow, from intuitive perspectives, the total amount of energy seems to remain constant to reach a certain level of excitation, but for case 2 this energy is spread during a longer period. It is also clear from Figure 4 how the algorithm is, for case 2, avoiding ventilation signals with big changes and thus avoids bang-bang behaviors, again due to the constraints in the energy usage.

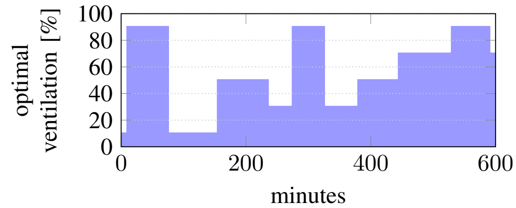

Fig. 4. The obtained optimal ventilation signal for the case $\beta = 0.2$.

B. Parameters and occupancy estimation

We then use 3 different datasets for estimating the parameters using the PEM approach described in Section II-B. We use these 3 different estimated parameters to find 3 different estimates of the occupancy signal using the regularized deconvolution method proposed in [9] on an other set of synthetic data (to do not run estimation of the occupancy steps on the same set used for estimating the parameters).

The 3 different datasets are then the two described in Section V-A and the one for which the ventilation signal is $\boldsymbol{u}^*$ as in (20), i.e., for which the ventilation signal has not been designed with the purpose of estimating occupancy levels but is rather the normal ventilation signal that one may have in the room (for completeness, the number of samples in this dataset is equal to the number of samples in case 2).

To assess the effectiveness of the overall input design procedure we then run 50 Monte Carlo (MC) simulations for different realizations of the measurement noise, and compute for each MC run and for each set of estimated parameters the value of the application function defined in (11). The results, shown in Figure 5, indicated that for both case 1 and case 2 the accuracy of the consequent occupancy estimator is comparable. For the dataset obtained without experiment design steps, instead, the performances of the occupancy estimator decay drastically of approximately a factor of 4 in terms of the MSE (see [9]) of the estimators. This thus indicates how important is to perform experiment design.

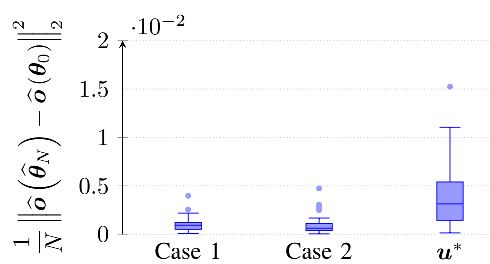

Fig. 5. Estimation errors of occupancy estimators trained in the datasets obtained in cases 1 and 2 in Section V-A and using $\boldsymbol{u}^*$, i.e., without applying experiment design concepts.

VI. Conclusions

We studied how to optimally ventilate a room during the collection of a training dataset to be used for identifying the effects of room occupancy on the CO_2 concentrations in the room so that the generated datasets are informative enough for designing occupancy estimation algorithms.

In other words, we developed an application-oriented input design approach for designing the ventilation signal dedicated to the problem of occupancy estimation. The method not only guarantees a certain accuracy in the estimated occupancy levels when using the collected data, but also tries to minimize the experiment time and effort, and account for potential restrictions on the CO_2 levels and the ventilation signal (due, e.g., to requirements on indoor air quality and actuators limitations). Finally, the problem is formulated as a nonlinear and nonconvex optimization problem that can be solved using a time recursive strategy. This eventually makes it possible to employ exhaustive search methods, and makes the overall scheme implementable in standard computers.

In-silico analyses of the results of the scheme confirm several intuitions: if the user does not limit the amount of excitation that the input design algorithm can use, the resulting ventilation signal becomes a bang-bang signal. If instead the user puts a limit on how much energy shall be used per time unit by the input, the computed ventilation signal becomes smoother and (at the same time) the experimental time becomes longer. Interestingly, though, the integral of the amount of energy spent by the system seems to remain practically constant. Finally, as expected, using input design methodologies lead to consequent occupancy estimators that perform approximately 4 times better in terms of the MSE.

Our future efforts are implementing and testing the strategy in real settings, and performing robust input designs.

References

[1] D. Sturzenegger, F. Oldewurtel, and M. Morari, "Importance of Long-Term Occupancy Information - A Validation with Real Occupancy Data," in *Clima - RHEVA World Congress*, 2013.

[2] T. A. Nguyen and M. Aiello, "Energy intelligent buildings based on user activity: A survey," *Energy & Buildings*, vol. 56, pp. 244–257, 2013.

[3] M. Gruber, A. Trüschel, and J.-O. Dalenbäck, "CO2 sensors for occupancy estimations: Potential in building automation applications," *Energy and Buildings*, vol. 84, pp. 548–556, Dec. 2014.

[4] J. Brooks, S. Goyal, R. Subramany, Y. Lin, T. Middelkoop, L. Arpan, L. Carloni, and P. Barooah, "An experimental investigation of occupancy-based control of commercial building climate," in *IEEE Conference on Decision and Control*, 2014.

[5] G. Liu, A. Dasu, and J. Zhang, "Review of Literature on Terminal Box Control, Occupancy Sensing Technology and Multi-zone Demand Control Ventilation (DCV)," U.S. Department of Energy, Tech. Rep. March, 2012.

[6] T. Labeodan, W. Zeiler, G. Boxem, and Y. Zhao, "Occupancy Measurement In Commercial Office Buildings For Demand-Driven Control Applications - A Survey And Detection System Evaluation," *Energy and Buildings*, vol. 93, pp. 303–3014, 2015.

[7] D. Calì, P. Matthes, K. Huchtemann, R. Streblow, and D. Müller, "CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings," *Building and Environment*, vol. 86, pp. 39–49, 2015.

[8] S. Wang, J. Burnett, and H. Chong, "Experimental Validation of CO2-Based Occupancy Detection for Demand-Controlled Ventilation," *Indoor and Built Environment*, vol. 8, no. 6, pp. 377–391, Nov. 1999.

[9] A. Ebadat, G. Bottegal, M. Molinari, D. Varagnolo, B. Wahlberg, H. Hjalmarsson, and K. H. Johansson, "Multi-room occupancy estimation through adaptive gray-box models," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 3705–3711.

[10] 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 *IEEE Conference on Decision and Control*, 2009, pp. 1494–1500.

[11] Z. Yang, N. Li, B. Becerik-Gerber, and M. Orosz, "A non-intrusive occupancy monitoring system for demand driven HVAC operations," in *ASCE Construction Research Congress*, 2012, pp. 2001–2010.

[12] B. Ai, Z. Fan, and R. X. Gao, "Occupancy Estimation for Smart Buildings by an Auto-Regressive Hidden Markov Model," in *American Control Conference*, 2014, pp. 2234–2239.

[13] K. P. Lam, M. Höynck, B. Dong, B. Andrews, Y.-s. Chiou, D. Benitez, and J. Choi, "Occupancy detection through an extensive environmental sensor network in an open-plan office building," in *IBPSA Conference*, 2009, pp. 1452–1459.

[14] B. Dong, B. Andrews, K. P. Lam, M. Höynck, R. Zhang, Y.-S. Chiou, and D. Benitez, "An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network," *Energy and Buildings*, vol. 42, no. 7, pp. 1038–1046, July 2010.

[15] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy Detection from Electricity Consumption Data," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys'13*. Rome, Italy: ACM Press, 2013, pp. 1–8.

[16] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and co_2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28 – 39, 2016.

[17] G. Goodwin and R. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*. New York: Academic Press, 1977.

[18] H. Hjalmarsson, "From experiment design to closed-loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, March 2005.

[19] X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, October 2006.

[20] A. Ebadat, G. Bottegal, D. Varagnolo, B. Wahlberg, and K. H. Johansson, "Regularized deconvolution-based approaches for estimating room occupancies," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1157–1168, 2015.

[21] H. A. Aglan, "Predictive model for CO_2 generation and decay in building envelopes," *Journal of applied physics*, vol. 93, no. 2, pp. 1287–1290, 2003.

[22] L. Ljung, *System Identification - Theory for the user*, 2nd ed. New Jersey: Prentice Hall, 1999.

[23] L. Lennart and P. E. Caines, "Asymptotic normality of prediction error estimators for approximate system models," *Stochastics*, vol. 3, no. 1-4, pp. 29–46, 1980.

[24] H. Hjalmarsson, "System Identification of Complex and structured systems," *European Journal of Control*, vol. 15, no. 3, pp. 275–310, 2009.

[25] C. R. Rojas, J. Welsh, G. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, pp. 993–1008, 2007.

[26] P. E. Valenzuela Pacheco, "On risk-coherent input design and bayesian methods for nonlinear system identification," Ph.D. dissertation, KTH Royal Institute of Technology, 2016.

[27] L. Gerencser, H. Hjalmarsson, and J. Mårtensson, "Identification of arx systems with non-stationary inputs - asymptotic analysis with application to adaptive input design," *Automatica*, vol. 45, no. 3, pp. 623–633, March 2009.

[28] C. R. Rojas, D. Katselis, H. Hjalmarsson, R. Hildebrand, and M. Bengtsson, "Chance constrained input design," in *IEEE Conference on Decision and Control and European Control Conference*, no. 5. IEEE, december 2011, pp. 2957–2962.

[29] C. A. Larsson, C. Rojas, and H. Hjalmarsson, "MPC oriented experiment design," in *18th IFAC World Congress*, no. 1986, Milano (Italy) August, 2011, pp. 9966–9971.

[30] G. C. Calafiore and M. C. Campi, "The Scenario Approach to Robust Control Design," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742–753, may 2006.

[31] B. Wahlberg, H. Hjalmarsson, and M. Annergren, "On Optimal Input Design in System Identification for Control," in *49th IEEE Conference on Decision and Control*, Atlanta, GA, USA, 2010.

[32] E. Zacekova, S. Privara, Z. Vana, J. Cigler, and L. Ferkl, "Dual control approach for zone model predictive control," in *Control Conference (ECC), 2013 European*. IEEE, 2013, pp. 1398–1403.