Energy and Buildings 126 (2016) 415–423

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Using synthetic population data for prospective modeling of occupant behavior during design

Clinton. J. Andrews*, MaryAnn Sorensen Allacci, Jennifer Senick, Handi Chandra Putra, Ioanna Tsoulou

Center for Green Building, Edward J. Bloustein School of Planning and Public Policy, Rutgers University, New Brunswick, NJ, USA

ARTICLE INFO

Article history:
Received 13 December 2015
Received in revised form 15 May 2016
Accepted 17 May 2016
Available online 18 May 2016

Keywords:
Occupant behavior
Energy
Simulation
Synthetic data

ABSTRACT

This paper addresses the challenge of incorporating occupant behavior into building performance simulation models used during the design process—that is, before the actual occupants are known. It proposes the use of synthetic population data, an approach that is novel in building performance modeling although common in urban planning and public health. A simpler approach embodied in the ASHRAE *Fundamentals* volume is to report standard distributions of values for behavioral variables, assuming that parameters vary independently of one another when in fact many co-vary or are interdependent. An alternative approach calibrates models of occupant behavior against actual occupants in specific existing buildings, but this raises questions of transferability. Needed is a database of "generic" occupants that designers can use prospectively during the design process. This paper documents a process of combining disparate field studies of commercial buildings into a larger occupant behavior database and generating a statistically similar synthetic data set that can be shared without compromising confidentiality requirements associated with field studies. The synthetic data set successfully incorporates much of the covariance structure of the underlying field data and supports multivariate modeling. Its scope and structure necessarily serve the needs of the associated modeling framework. Cooperative and systematic sharing of data by field researchers is crucial for building large enough data sets to serve as a behaviorally-robust basis for building design.

1. Introduction

Design choices directly determine many aspects of building performance, but occupant responses to comfort conditions they experience can also influence energy performance [1]. Occupants can adjust thermostats, open windows, or install fans or heaters, and building operators may monitor occupant concerns and dynamically respond to their behaviors and preferences [2–4]. Hence there is value in understanding potential occupant influences early in the design process, before occupancy actually occurs.

Building performance modelers by necessity often ignore aspects of occupant behavior by assuming fixed comfort targets and ignoring "unregulated" energy loads. This is the standard practice even though researchers have shown that occupants influence building performance by their choice of setpoints, schedules, and adaptive behaviors, many of which are heterogeneous, often habitual, and sometimes maladaptive [5,6]. Parsimonious, accurate and reliable representations of expected adaptive behaviors that respond to comfort conditions experienced in new buildings have not been available to most practitioners [7]. This paper presents one way to bring insights from occupant behavior research to the building design process that strives to balance parsimony and accuracy.

Data collected about occupant behavior inevitably comes from existing buildings and experimental setups, whereas designers must build new buildings whose occupants are not yet known. Some behavior is context-dependent, so there is a need to study how transferable the knowledge gained from retrospective study of occupant behavior is to new building contexts. This paper identifies four broad transfer approaches discussed in greater detail next: (1) develop standard distributions for design guidance; (2) calibrate building performance models against existing occupied buildings, then apply the calibrated building performance model elsewhere; (3) calibrate separate occupant behavior and building performance models in co-simulation against an existing building, then apply the calibrated occupant behavior model elsewhere; and (4) develop a

* Corresponding author.
E-mail address: CJA1@RUTGERS.EDU (Clinton.J. Andrews).

http://dx.doi.org/10.1016/j.enbuild.2016.05.049
0378-7788/

representative occupant behavior data set for use in co-simulations that link occupant behavior and building performance models.

1.1. Develop standard distributions

Large data sets incorporating thousands of occupants in hundreds of buildings underlie standard distributions of thermal comfort preferences summarized in the ASHRAE *Fundamentals* volume [8] and adopted in ASHRAE *Standard 55* [9] and the associated primary literature, e.g., Ref. [10]. Occupants' predicted mean votes (PMV) on the ASHRAE 7-point thermal comfort scale, which ranges from cold (-3) to hot (3), depend strongly on their metabolic rate and clothing type, plus ambient air temperature, mean radiant temperature, air velocity, and relative humidity. PMV, in turn, drives the predicted percentage of dissatisfied occupants (PPD). ASHRAE *Standard 55* [9] focuses its guidance on the central tendency in this distribution, recommending that designers seek to satisfy the 90% of occupants whose PMV lies between −0.5 and +0.5 on the thermal comfort scale. Designers may assume fixed indoor air temperature targets for use in subsequent analysis and equipment sizing calculations.

Complicating contextual factors may cause the standard thermal comfort distribution to shift between summer and winter, and between buildings with mechanical and natural ventilation [11]. Indeed, "because of the large interpersonal variability in thermal requirements, some occupants in any uniformly conditioned environment will be too warm at the same time as others are too cool" [8]. A comparison of the 1989 and 2013 versions of the ASHRAE *Fundamentals* thermal comfort chapter shows that the older guidance focused on the central tendency in the distribution of occupant perceptions and responses, but the more recent version focuses also on the distribution of perceptions and behaviors around that central tendency, which opens up new design possibilities such as for personal environmental control systems [8,12]. On this basis, designers may assume a *distribution* of desired indoor air temperatures when performing analysis and selecting equipment.

The standard distribution approach is transferable in the sense that it has been disseminated by ASHRAE [8] and is easy to use. However, it fails to capture the dynamic effects of feedbacks including adaptive responses by occupants to changing building conditions. Simulation modeling is necessary to add the dynamics.

1.2. Calibrating models using occupied buildings

Most building performance models simulate a dynamic system using deterministic equations driven by time-varying physical and behavioral parameters [13]. In an extension of the standard distribution approach discussed previously, guidance manuals such as ANSI/ASHRAE/IES *Standard 90.1* Appendix G [14] specify typical occupancy schedules, diversity factors and other standardized behavioral assumptions to include in simulations of new buildings. However, modelers find that the resulting simulations may not match reality when calibrated against existing buildings [15].

Standardized calibration approaches have therefore emerged, with most relying on manual, iterative adjustments of parameters for a few key variables [16]. A growing literature recommends setting goodness-of-fit criteria in advance, gathering detailed data over time and by zone, seeking independent measurement of weather-related and internal loads, tuning weather-dependent variables first, and expecting more experienced modelers to fare better than those with less experience [17–22]. A complementary literature recommends statistical strategies for managing this underdetermined optimization problem by using systematic searches or heuristic strategies to identify the most influential variables, using penalty functions to reduce overall error, and taking spot measurements to reduce key uncertainties [23,24,16]. A consistent finding is that statistical noise due to occupancy patterns and occupant behavior limits the accuracy of calibration efforts [25,26]. Most of these factors limit the transferability of calibrated models from existing to future buildings. Instead, it seems that the experience of the modeler is the most transferable element.

1.3. Co-simulating occupant behavior & building performance

Social and behavioral scientists have a long tradition of modeling human behavior [27,28], and computer scientists have developed a parallel tradition relying on a procedural rather than correlative framing of behaviors, e.g., Refs. [29–31]. Recently a new class of building energy performance simulation models focusing on occupant behavior has emerged that favors probabilistic, Markov process, and agent-based modeling approaches to represent interactions between building systems and occupants, e.g., Refs. [32–35]. Of particular interest are models that co-simulate occupant behavior and building performance by dynamically coupling behavioral models with standard design tools such as *EnergyPlus* using interface ontologies such as *obXML* [36,37].

Calibration of a coupled occupant behavior and building performance modeling system is an elaborate and interactive process that involves collecting the usual detailed data on building geometry and energy consumption at both the component and building-wide levels, as well as detailed data on occupancy patterns, occupant sensations and perceptions, and their adaptive responses to conditions in the building [38]. As was the case with building performance models discussed previously, calibration efforts are typically manual, iterative and rely on heuristics and the modeler's expertise. The constraining factor is usually the limited amount of occupant behavior data that is available, even for a well-funded study of an existing building.

While detailed models of this type hold interest for researchers, not all have practical value, and often a more parsimonious, fit-for-purpose approach is warranted [7]. One parsimonious strategy is to present practitioners with a generic and hopefully representative occupant data set for use in modeling.

1.4. Representative occupant population

Most research in the occupant behavior domain involves small data sets. Some researchers pursue longitudinal studies of a small number of occupants, e.g., Ref. [36], others do cross-sectional studies of a larger number of occupants, e.g., Ref. [39]. The typical calibration approach combines top-down and bottom-up elements by codifying a theory of human behavior in equations and fitting the parameters of those equations. One approach, reflecting the current state of data, is to focus on the distribution of values for each modeling parameter separately, that is, assuming independent bivariate functional relationships. This allows the modeler to borrow parameters and standard distributions from others to apply in the model when local data are not available. The modeling framework then needs to incorporate logic showing how these functional relationships interact to yield multivariate covariance, as is shown in Ref. [36]. A second approach, which the current paper advances, is to focus even more ambitiously on a representative population of occupants in order to capture interactions and covariation within a defined set of contextual parameters. Such a data set would be particularly helpful for calibrating agent-based models of occupant behavior but it should also be useful in other modeling traditions.

It is easier to imagine a representative population of building occupants than to find one that is well documented. There are many potential variables of interest, such as age, gender, schedule (daily, weekly, seasonal), metabolism, clothing, multiple adaptive behaviors, and comfort perceptions, and many measured values are context-dependent [40]. Currently, there is much data on ther-