A physics-based domain adaptation framework for modeling and forecasting building energy systems

Data-Centric Engineering (2023), 4: e10
doi:10.1017/dce.2023.8

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

A physics-based domain adaptation framework for modeling and forecasting building energy systems

Zack Xuereb Conti[1,2], Ruchi Choudhary[1,2] and Luca Magri[3]

[1]Data-Centric Engineering Program, The Alan Turing Institute, London, United Kingdom
[2]Department of Engineering, University of Cambridge, Cambridge, United Kingdom
[3]Faculty of Engineering, Department of Aeronautics, Imperial College London, London, United Kingdom
Corresponding author: Zack Xuereb Conti; Email: zxuerebconti@turing.ac.uk

Received: 22 August 2022; **Revised:** 09 February 2023; **Accepted:** 31 March 2023

Keywords: Building energy forecasting; domain adaptation; dynamical systems; state-space models; subspace alignment

Abstract

State-of-the-art machine-learning-based models are a popular choice for modeling and forecasting energy behavior in buildings because given enough data, they are good at finding spatiotemporal patterns and structures even in scenarios where the complexity prohibits analytical descriptions. However, their architecture typically does not hold physical correspondence to mechanistic structures linked with governing physical phenomena. As a result, their ability to successfully generalize for unobserved timesteps depends on the representativeness of the dynamics underlying the observed system in the data, which is difficult to guarantee in real-world engineering problems such as control and energy management in digital twins. In response, we present a framework that combines lumped-parameter models in the form of linear time-invariant (LTI) state-space models (SSMs) with unsupervised reduced-order modeling in a subspace-based domain adaptation (SDA) approach, which is a type of transfer-learning (TL) technique. Traditionally, SDA is adopted for exploiting labeled data from one domain to predict in a different but related target domain for which labeled data is limited. We introduced a novel SDA approach where instead of labeled data, we leverage the geometric structure of the LTI SSM governed by well-known heat transfer ordinary differential equations to forecast for unobserved timesteps beyond available measurement data by geometrically aligning the physics-derived and data-derived embedded subspaces closer together. In this initial exploration, we evaluate the physics-based SDA framework on a demonstrative heat conduction scenario by varying the thermophysical properties of the source and target systems to demonstrate the transferability of mechanistic models from physics to observed measurement data.

Impact Statement

This paper addresses generalization limitations with traditional data-driven modeling methods applied to building energy forecasting by introducing a transfer-learning approach that combines physics-based lumped-parameter models in the form of linear state-space models (SSMs) and unsupervised reduced-order modeling methods. Instead of learning black-box models whose generalizability to forecast for unobserved timesteps depends wholly on the representativeness of underlying dynamics in the data, our aim is to leverage the governing structure of low-rank SSMs in a domain adaptation framework. SSMs are well established for building energy forecasting and control purposes and are straightforward to derive from well-known energy transfer ordinary differential equations.

1. Introduction

This paper addresses generalization limitations of traditional data-driven methods when applied to building energy modeling and forecasting. We present a novel transfer-learning (TL) framework that combines physics-based models widely used in the energy community for model predictive control (MPC), with domain adaptation (DA), which is a TL technique typically adopted to leverage pre-trained models for prediction across different but related tasks.

1.1. Building energy modeling and forecasting

The urgency of an ongoing climate crisis in addition to a rising demand for thermal human comfort highlights the need to mitigate energy demand in buildings. Over the last decade, building energy control and operation strategies coupled with innovations in smart grid infrastructures at the urban scale, and controllable energy systems such as heating, ventilation and air-conditioning (HVAC), battery storage, and renewable energy generation at localized building scale, has emerged as a significant approach toward manipulating the energy demand in daily post occupancy operations. On the other hand, advanced energy analysis and optimization methods in preoccupancy stages assist the mitigation of critical design factors influencing both passive and operative energy demand generation. More recently, digital twins that couple energy models with measurement data provide a framework to optimize energy systems of buildings in real time.

Building energy modeling and forecasting plays a vital role in building energy mitigation during both pre and post occupancy stages. A significant volume of past and recent literature (Li and Wen, 2014a; Tardioli et al., 2015; Bourdeau et al., 2019) categorize techniques for building energy modeling and forecasting into three main approaches: *physics-based*, *data-driven*, and *hybrid* models. Physics-based models, also known as white-box models (Tardioli et al., 2015), are derived directly from mechanistic knowledge of physics principles. Data-driven models, also known as black-box models (ASHRAE, 2009; Tardioli et al., 2015), rely on timeseries data and machine-learning (ML) algorithms to accurately learn input–output relationships; these include classic linear regression and more advanced data-driven methods such as random forest (RF) (Wang et al., 2018), support vector machine regression (SVR) (Shao et al., 2020) and long short-term memory (LSTM) (Sehovac et al., 2019; Wang et al., 2019; Somu et al., 2020). Hybrid models, also referred to as grey-box models (Tardioli et al., 2015), are a combination of both physics-based and data-driven models that use simplified physical descriptions of building energy systems via low-rank models and subsequently use data to infer the model parameters via system identification methods; these include resistance–capacitance (RC) thermal networks and state-space models (SSMs) (Candanedo et al., 2013; Amara et al., 2015; Fateh et al., 2019).

While physics-based models are fully interpretable, they often become infeasible for complex energy systems such as largescale building with several thermal zones because they require large numbers of parameters to be specified, coupled with a computational expense to solve. In such cases, it is typical to either simplify the energy system via hybrid modeling (Goyal and Barooah, 2011) or learn an input–output black-box surrogate model via data-driven modeling (Rätz et al., 2019). The latter has become a widely popular approach for building energy modeling and forecasting due to the recent emergence of ML techniques where given enough data, they are good at finding spatiotemporal patterns and structures even in scenarios where the complexity prohibits analytical descriptions (Willard et al., 2022). In the building energy community, vast literature has demonstrated that ML-based energy modeling methods outperform classic statistical methods in terms of accurately capturing spatiotemporal structures from data for short-term forecasting (Wang et al., 2018, 2019; Sehovac et al., 2019; Shao et al., 2020; Somu et al., 2020). However, ML-based models are prone to caveats that often impede their application to real-world engineering applications such as real-time monitoring in digital twins (Karpatne et al., 2017). In this paper, we identify these caveats as (a) model-interpretability, (b) model-generalizability, and (c) data-dependency.

The underlying structure of ML-based models such as for example, deep neural networks, typically take the form of a so-called multi-layer perceptron; a convoluted graph network connecting inputs to

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

outputs. The graph does not hold physical correspondence to mechanistic structures linked with governing phenomena thus render such models difficult to interpret. The network topology and corresponding hyperparameters are typically determined from data using black-box identification techniques. Consequently, the ability for the black-box model to generalize for unseen inputs is heavily influenced by how well the governing behavior of the true system underlying the data is represented in the dataset used for training (not simply on the size of the dataset, as is commonly assumed). If for example, certain occurrences in the dataset are observed more frequently than is characteristic of the true system's dynamical behavior then, training on such data introduces bias to the model-learning, resulting in poor predictive generalization (Kouw and Loog, 2018). Since there is no easy way to determine how complete a governing behavior is represented in a dataset or how to control the "representativeness"; black-box models rely on the size of the dataset to secure generalization robustness, in a hit a mis approach. Thus, we can argue that improving the generalizability of typical black-box models for energy forecasting may result in a data-intensive task, which may not be suitable for applied scenarios where for example, obtaining measurement data is challenging or not feasible.

While all caveats need to be addressed, in this paper we focus on addressing *generalizability.* Specifically, we aim to improve the generalization of building energy models such that we can forecast for unobserved timesteps given building measurement timeseries data. We present an approach that combines physics-derived lumped-parameter models (LPMs) widely adopted in the building energy community for simplified thermal modeling, with DA which is a type of TL technique to leverage pretrained models for prediction across different but related tasks. In summary, our goal is to forecast for unobserved measurement data by leveraging the generalizability innate in the governing dynamics derived from well-known ordinary differential equations (ODEs) of heat transfer, mechanistically represented as LPMs, even if the latter describes the observed system approximately.

1.2. Lumped-parameter models for building thermal modeling

Thermal modeling of real-world building scenarios may involve high-order intricacies. In scenarios where real-time computation is of essence, such as MPC in digital twins, or iterative computer-based optimization, it becomes infeasible to model and solve such models analytically nor numerically. Instead, LPMs are a popular alternative because they balance out computation time and accuracy while retain mechanistic integrity of the physics to a reasonable degree. Such model simplification methods are also referred to as low-rank models. A popular LPM class of thermal modeling in buildings are RC networks that use an electric circuit analogy to represent the principal energy flow and energy transfer phenomena governing the energy behavior at varying scales; across thermal zones and building components whose behavior is influenced by thermal dynamics (Lin et al., 2012; Fayazbakhsh et al., 2015; Koeln et al., 2018), and across urban scales (Bueno et al., 2012). An implementation example of an RC network can be found in the demonstrative example in Section 2. When representing thermal dynamics of a building as an RC network, the capacitors represent thermal capacitance of the air within a zone or the material of a building component, while the resistors represent thermal resistance of the medium between adjacent thermal capacitances (Goyal and Barooah, 2012; Amara et al., 2015). Once assembled, the network of resistors and capacitors reflect an adjacency network of thermal zones and components via interconnecting energy flow paths. The heat transfer dynamics at each capacitance node can be described by ODEs, whose parameters are either derived from well-known laws of physics or inferred from data through system identification techniques.

In building control applications such as MPC, RC networks are typically formulated in a more mathematically standardized formulation where the coupled system of ODEs in the RC network are written as a linear time-invariant (LTI) SSM, which is a more compact format for representing the mechanistic structure of the transient thermal dynamics that, for example, relates the control signals to the space temperatures and humidity of each zone. Various applications of LTI SSMs to building energy modeling have shown that linear models sufficiently balance out prediction accuracy and model simplicity, as required for MPC (Candanedo et al., 2013; Li & Wen, 2014b; Picard et al., 2015). Further

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

literature such as Goyal and Barooah (2012) present a reduced-order modeling (ROM) technique catered for reducing the order of energy SSMs with large number of states, while maintaining a reasonable prediction accuracy. In the case of nonlinearities present in the building energy model due to for example, longwave radiation exchange, absorption of incident solar radiation at the facade, or convective heat transfer, Picard et al. (2015) suggest linearization techniques by retaining the nonlinear terms in their state-space formulation.

Despite being linear and approximate, we adopt LTI SSMs in our framework because of their great convenience for control design and estimation, which is justified further by their popularity in the building energy community and by their reasonable robustness in preserving the important dynamics, as illustrated in the above literature. Unlike purely data-driven black-box models, LTI SSMs facilitate access to the mechanistic structure responsible for governing the observed dynamics, even when the physics only describes the dynamics of the system at hand, approximately. It is the innate generalizability thanks to their mechanistic structure, which we aim to leverage using TL.

1.3. Transfer learning

Transfer learning breaks the notion of traditional ML where the training and testing data must come from the same feature space and similar distributions. In fact, in scenarios, where training data are limited, TL strategies can be useful to adapt a trained model from a different but related task. TL methods can be categorized depending on the availability of labeled data in the source and target domains, as follows: (a) inductive TL, where both source and target labeled data are available; (b) transductive TL, where labeled data are only available at the source domain but limited to no labeled data are available at the target domain; and c) unsupervised TL is similar to inductive learning but assumes no labeled data available. A full description of each of these categories, and further subcategories can be found in an extended survey on TL techniques (Pan and Yang, 2010).

In the building energy community, the application of TL is recent. Most applications aim to overcome the dependency on historical data by transferring or adapting data-driven models trained for buildings with available historical measurement data (labeled data), to forecast energy in physically similar buildings whose historical data is limited (some labeled data) or completely unavailable (unlabeled). For example, Fang et al. (2021) address short-term energy prediction for limited historical data by introducing a hybrid deep TL strategy where LSTM-based extractor is used to infer spatiotemporal features across source and target buildings, respectively. Subsequently a domain-invariant feature space is learned to bring closer the two domains using a domain adaptation neural network (DANN) and thus, the model trained on source building data can be leveraged to predict energy in a target building. Similarly, Ribiero et al. (2018) introduce a novel TL method based on timeseries multi-feature regression with seasonal and trend adjustments. Their approach can be used for buildings with small data by leveraging data from similar buildings with different distributions and seasonal profiles. Their work was validated using a case study involving energy prediction for a school using additional data from other schools. With a similar goal, Gao et al. (2020) present a TL approach combining a sequence to sequence (seq2seq) model with a convolutional neural network (CNN), resulting in significant accuracy improvements over the use of standard LSTM given only some data in the target building. Their cross-building approach also collects data from similar buildings in a TL framework and is validated by predicting energy for three government buildings across two cities in China. Mocanu et al. (2016) present a TL approach to predict the energy consumption at the building energy level in a smart grid context without the need of labeled historical data from the target building by combining a deep belief network (DBN) for feature extraction, with an extended reinforcement learning approach for knowledge transfer (KT) between buildings. The DBN estimates continuous states, which are then incorporated into the reinforcement learning methods in a continuous lower-dimensional representation of the energy consumption. The outcomes show that their model generalizes for varying time horizons and resolutions using real data thanks to the generalization power of deep belief networks.

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

In our approach, we adopt a type of transductive TL technique called *subspace-based domain adaptation* (SDA), which differs from other TL methods by assuming the existence of a domain-invariant feature space between the source and target domains. Using a transductive approach, we leverage the facility of "labeled data" using the physics-based LTI SSM, by adapting the model to forecast for unobserved timesteps (unlabeled data) beyond measurement data. SDA is also referred to as *KT* in literature. The authors are not aware of previous applications of SDA to building energy modeling. More generally, we believe DA techniques hold significant untapped benefits for building energy modeling applications, including data-efficient calibration of models and cross-building generalization, to name a few.

SDA is a type of TL approach that involves embedding/projecting source and target data into lower-dimensional vector spaces called *subspaces.* The source and target subspace are defined by the eigenvectors obtained from the ROM methods, respectively. Subsequently SDA seeks for a geometric transformation that brings the subspaces closer together in the form of a mutual subspace (Pan et al., 2011). Projecting the respective data onto a lower-dimensional subspace helps to reduce the difference between distributions between domains as much as possible while preserving important properties of the original data, such as geometric and statistical properties.

SDA is widely studied in literature with advantageous applications in fields such as computer vision (Baktashmotlagh et al., 2013; Fernando et al., 2013), natural language processing (Blitzer et al., 2006), and so forth. Most research in SDA goes toward improving the lower-dimensional embedding and/or the subspace alignment (SA) strategies with the ultimate scope of minimizing the distance between source and target. For example, Pan et al. (2011) present an SDA approach to find a mutual subspace using transfer component analysis (TCA), which combines feature embedding via principal component analysis (PCA) with maximum mean discrepancy (MMD) to find an optimal shared subspace in which the distribution between source and target domain datasets is minimized. In their work, they discuss both unsupervised and semi-supervised feature extraction approaches, to dramatically reduce distance between domain distributions by projecting data onto the learned transfer components. In another approach, Huang et al. (2012) adopt canonical correlation analysis (CCA) to infer a correlation subspace as a mutual representation of action models captured by different cameras for cross-view action/gesture recognition. Elhadji-Ille-Gado et al. (2017) introduce an SDA framework where source and target domains are embedded, respectively, into subspaces described by eigenvector matrices via an approximated singular value decomposition (SVD) method. Subsequently, the source subspace eigenvectors are reoriented to align closer towards the target eigenvectors via SA, resulting in a target-aligned source subspace. In a similar approach, Fernando et al. (2013 and Sebban et al. (2014) present a closed form method leading to a mapping function (transformation matrix) that aligns source subspace with target subspace vector bases, respectively derived via PCA. By aligning the bases of the subspaces, their approach is global.

1.4. Proposed approach: A physics-based domain adaptation framework

Drawing inspiration from the above approaches, in this paper we present an SDA framework similar to Fernando et al. (2013) and Sebban et al. (2014). However, instead of leveraging labeled data from a source subspace, we leverage the structure of LPMs widely adopted for thermal modeling of buildings and building components, as discussed in Section 1.2. More specifically, instead of a source subspace derived from labeled data, we derive the subspace directly from the eigenvectors governing the structure of the LTI SSM whose parameters are derived analytically from well-known heat transfer ODEs. On the other hand, we describe the target subspace by the eigenvectors inferred from the measurement data via unsupervised ROM methods such as principal orthogonal decomposition (POD). Once both source and target subspaces are derived, we embed LTI SSM-simulated data (source) and observed measurement data (target) onto their respective lower-dimensional subspaces defined by the respective eigenvectors, and subsequently seek for a geometric transformation that brings the source subspace closer toward the target subspace. In dynamical systems theory, eigenvectors hold direct correspondence with the structure governing the dynamical behavior. Thus, projecting source and target data onto their respective subspaces

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

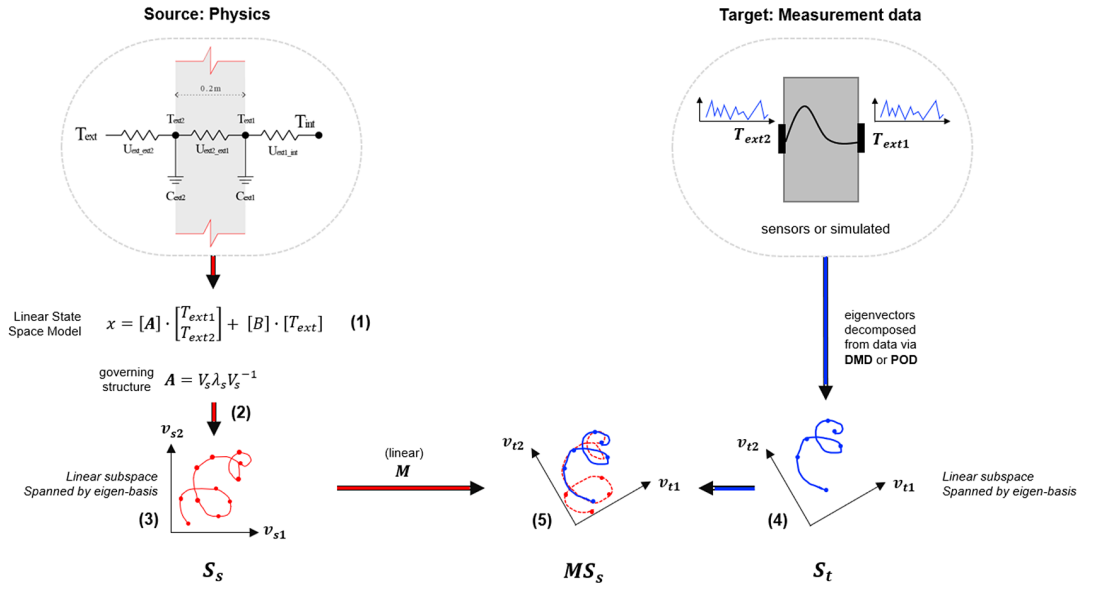

Figure 1. *Overall subspace-alignment-based domain adaptation (SDA) workflow.*

spanned by their respective eigenvectors, reduces the distance between the respective systems while facilitates a more data-efficient transfer. In Section 5, we illustrate how the SDA approach can be used to forecast unobserved timesteps beyond available measurement data using both near and distant physics-based approximations of the observed system.

In more detail, our physics-based SDA approach via SA can be organized into the following steps (see Figure 1): (1) specify the physics-based model, (2) center the data, (3) infer the source subspace, (4) infer the target subspace, (5) learn the transformation map to align the source subspace with the target subspace, and finally (6) project the source data onto the target-aligned source subspace to forecast the target domain.

1.5. Outline

In Section 2, we introduce a demonstrative scenario concerning a simple thermal system that will be carried through the paper. In Section 3, we describe the specification of the physics-based SSM (step 1). Subsequently, in Section 4, we describe the proposed subspace-based alignment (steps 2 to 5). Finally, in Section 5, we apply the proposed approach to the demonstrative scenario and discuss the results.

2. Demonstrative Example

We consider a demonstrative building energy system consisting of the thermal conductance through an external wall (outlined in red in Figure 2a) of a single thermal zone. The principal energy dynamics governing the energy transfer through the external wall are represented by the RC network in Figure 2b, where T_{ext}, T_{ext2}, T_{ext1}, and T_{int}, represent the external ambient temperature, the external surface temperature, the internal surface temperature, and the internal ambient temperature, respectively. In thermal RC networks, the spatial location of the nodes is typically selected with strategic intent for the identification of hidden thermal states of interest that may not be observed or measured directly. For example, in the RC representation of walls, the node locations are typically assigned between the discretized layers composing the construction of the wall to represent the state of interaction between adjacent thermal layers thus, capture the hidden states that play a role in governing the observable dynamic behavior. In our demonstrative example in Figure 2, nodes T_{ext2} and T_{ext1} were selected with the intention to capture the conductive transfer behavior governed by the thermophysical properties of the

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

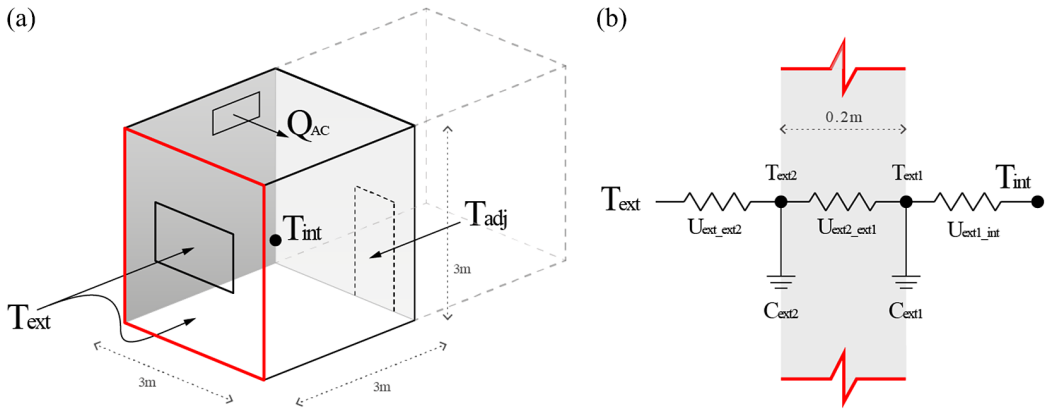

Figure 2. *(a) Thermal zone for context. (b) Thermal RC network model of one-dimensional energy transfer through an external wall.*

wall and represented by the thermal resistor where $R_{\text{ext2\_ext1}} = 1/U_{\text{ext2\_ext1}} R_{\text{ext2\_ext1}}$. The external wall separates two volumes of air where, the thermal resistors R_{ext_ext2} represents the convective heat transfer between external air and external wall surface while and R_{ext1_int} represents the convective heat transfer between the inner wall surface and internal room air. Note, that we assume a one-dimensional heat flow through the wall while we ignore the influence of solar radiation acting on the outer surface of the wall. The lumped RC network in Figure 2b is commonly labeled as the 3R2C configuration (three resistors and two capacitors) and is a well-established and validated simplified model for simulating energy behavior of building envelopes and roofs for MPC applications in building management systems (Haves et al., 1998).

We can model the system of nodes algebraically using the general heat balance equation, given by

$$\frac{dT_j}{dt} = \frac{1}{C_j}\left[\sum \frac{1}{R_k}(T_k - T_j) + \sum Q_j\right], \tag{1}$$

where, j and k represent two adjacent thermal nodes; C_j represents the thermal storage capacity at node j; R represents the thermal resistance connecting j and k; Q represent heat gains from external sources such as lighting and equipment. For the simple wall scenario, we assume a homogenous 1D energy conduction through the wall while ignore heat gains Q from the zone adjacent to the inner surface of the wall. Using equation (1), we can derive the respective energy balance equations for each node $[T_{ext1}, T_{ext2}]$ in the 1D wall scenario as follows:

$$\frac{dT_{ext1}}{dt} = \frac{1}{C_{ext1}}\left[\frac{1}{R_{ext2_{ext1}}}(T_{ext1} - T_{ext2})\right], \tag{2}$$

$$\frac{dT_{ext2}}{dt} = \frac{1}{C_{ext2}}\left[\frac{1}{R_{ext2\_ext1}}(T_{ext1} - T_{ext2}) + \frac{1}{R_{ext\_ext2}}(T_{ext2} - T_{ext})\right]. \tag{3}$$

The R and C parameters can either be identified using various black-box system identification optimization methods (Madsen, 1995; Candanedo et al., 2013; Sarkar et al., 2021) from numerically-simulated or empirically-measured data or can be derived directly from first principles. Given our aim to leverage physics-derived knowledge of phenomenon, we derive R and C directly from first principles of heat storage capacitance, convective resistance, and conductive resistance using well-established mechanisms (equations (4)–(6)).

$$C = Cp\rho V, \tag{4}$$

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

$$R_{conv} = \frac{1}{hA,} \tag{5}$$

$$R_{cond} = \frac{kA}{w/2}. \tag{6}$$

Table 1

Table 1. *Assigned physical and thermodynamic properties for the 1D wall scenario.*

Genre	Property, notation	Value
Physical properties	Wall layer volume, V	1.8 m^3
	Wall thickness, w	0.2 m
Thermophysical brick properties	Brick conductivity, k	0.72 W/mK
	Brick density, ρ	1,920 kg/m^3
	Brick specific heat capacity, Cp	780 J/kgK
Convection coefficients	Indoor convection, h	8 W/m^2K
	Outdoor convection, h	25 W/m^2K
Thermophysical air properties	Air density, ρ	1.2 kg/m^3
	Air specific heat capacity, Cp	100 J/kgK

3. Specification of the Physics-Based Model (Step 1)

The first step in our SDA approach is the specification of the physics-based model. In contrast to standard SDA approaches discussed earlier, in our approach, we assume a physics-based LPM as the source domain. More specifically, we assume a physics-derived LTI SSM which offers a standardized representation of dynamical systems and has been shown to capture governing behavior of energy in buildings accurately even when parameters are lumped (Goyal et al., 2011; Goyal and Barooah, 2012). Furthermore, lumped SSMs for building thermal network representations can be relatively straightforward to implement.

3.1. LTI SSM

We can further represent the coupled ODEs in equations (2) and (3) in state-space form which is a more compact format to represent the mechanistic structure of the dynamics. SSMs are a standardized representation to describe the dynamics of any system which can be described by a set of ODEs or partial differential equations (PDEs) (Hendricks et al., 2008). In general, SSMs consist of two matrix equations: the state equation (equation (7)) describes how the inputs influence the states of the system while the output equation (equation (8)) describes how the states of the system and the external factors (inputs) directly influence the observed output of the system. In this paper, we focus on discrete-time LTI SSMs which have the following form:

$$x(t) = [A] \cdot x(t-1) + [B] \cdot u(t-1), \tag{7}$$

$$y(t) = [C] \cdot x(t-1) + [D] \cdot u(t-1). \tag{8}$$

LTI implies that the coefficient parameters do not vary with time. From here on, it is implied that SSM refers to LTI SSM. SSMs are written in terms of three vectors: a state vector x with n elements; an input vector u with p elements; and an output vector y with q elements. Such a state-space formulation is equivalent to the well-known finite difference method. We find that the SSM formulation provides a

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

compact and standardized format to preserve the interpretable structure of governing dynamics, with focus for KT purposes described in Section 3.

When representing RC thermal networks as SSMs, we can adopt the temperatures at the capacitance nodes in the RC network as the state variables in the state vector x, to maintain the same intuitive interpretation (Candanedo et al., 2013). In our example wall scenario, the temperature on the internal surface of the wall is our variable of interest and thus, we consider state T_{ext1} as the output variable y while T_{ext} as the control input u, as follows:

$$x = \begin{bmatrix} T_{ext1} \\ T_{ext2} \end{bmatrix},\ u = [T_{ext}],\ y = [T_{ext1}]. \tag{9}$$

These vectors are linked by four coefficient matrices, A $(n \times n)$, B $(n \times p)$, C $(q \times n)$, and D $(q \times p)$ that capture the dynamic temporal behavior between states x, inputs u, and output y. More specifically, A is referred to as the dynamics operator which contains the dynamical characteristics to advance the state matrix x from in equation (7) from $x(t)$ to $x(t+1)$, for time step dt, for any arbitrary $x(t)$. B determines how the influences from control signals, for example, the outdoor air temperature, affect the states, for example, the inner and outer wall surface temperatures. The zero values in B indicate that the outdoor temperature T_{ext} influences the internal wall surface temperature T_{ext1} by conduction through the wall, not directly. The coefficient values of A, B, C, D are determined directly using the heat balance equations at each node. Thus, the coefficients for the wall example are determined via equations (2) and (3) and as follows:

$$A = \begin{bmatrix} \dfrac{-U_{ext1,ext2}}{C_{ext1}} & \dfrac{U_{ext1,ext2}}{C_{ext1}} \\ \dfrac{U_{ext1,ext2}}{C_{ext2}} & \dfrac{-U_{ext1,ext2} - U_{ext2,ext}}{C_{ext2}} \end{bmatrix},\ B = \begin{bmatrix} 0 \\ \dfrac{U_{ext2ext}}{C_{ext2}} \end{bmatrix}, C = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, D = [0]. \tag{10}$$

3.2. Structure of the governing state space

A state-space representation allows for a more geometric understanding of dynamical systems. The operator A can be decomposed into its fundamental components by

$$AV = V\lambda, \tag{11}$$

where, V and λ are the eigenvectors v_1, v_2 and eigenvalues λ_1, λ_2 of describing the dynamics of A, respectively. The system A can then be solved for any timestep dt and initial condition $x(t_0)$ by solving the matrix exponential method as follows:

$$\Phi_A = e^{dtA} = Ve^{dt\lambda}V^{-1}, \tag{12}$$

$$\dot{x}(t) = e^{tA}x(t_0), \tag{13}$$

where Φ_A is referred to as the state transition matrix; the matrix that updates the state vector from one timestep to the next. The eigenvalues lie on the diagonal of the A matrix. Thus, when computing the matrix exponential, we are scaling/transforming the eigenvalues to a given timescale, dt.

We demonstrate the generalizability and interpretability of this physics-derived model using portrait analysis, which is a qualitative analysis method from dynamical system theory to visualize the structure and solutions of dynamical systems geometrically, as trajectories in a vector field. Using portrait analysis, we aim to illustrate how V and λ govern the global physical behavior of dynamics, in our case, thermal conduction through the façade wall of a building.

In portrait analysis, instead of viewing the state variables (T_{ext1}, T_{ext2}) independently as functions of time (as a timeseries in Figure 3a), we can view their values as coordinates of a point in a vector space bound by T_{ext1}, T_{ext2} called the *state space* (Figure 3b). Thus, a point in the state space represents the complete state of the system at any time t. As the system evolves over time, the point will trace out a curve in the state space, referred to as a *trajectory.* Figure 3b illustrates the traced trajectory in the state space for

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

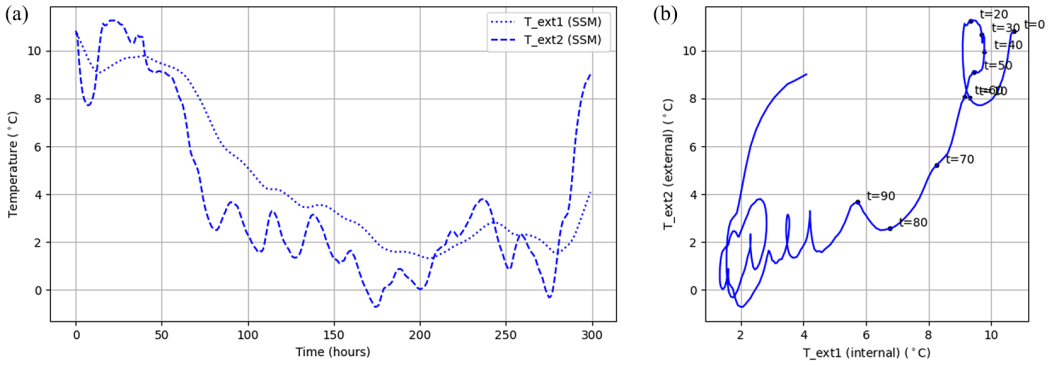

Figure 3. *(a) SSM-generated timeseries measurement data plotted as a function of time. (b) SSM-generated timeseries measurement data plotted in the state space.*

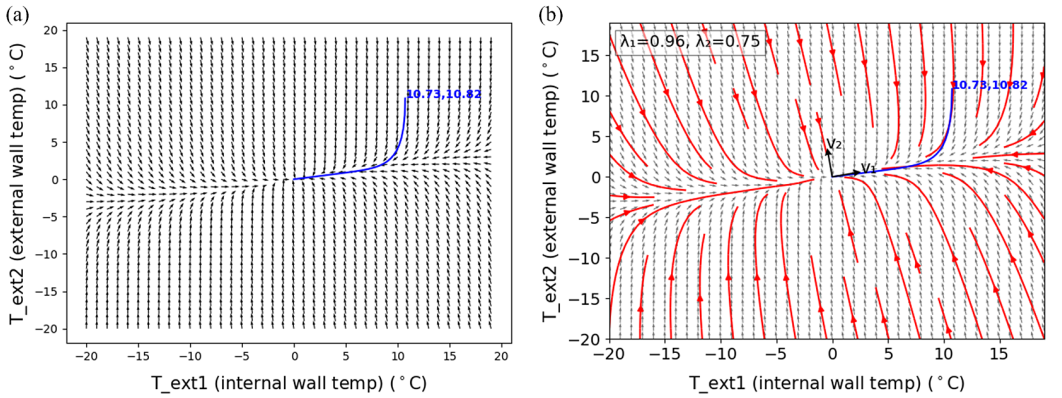

Figure 4. *(a) State-space vector field. (b) Phase portrait of the state space.*

300 hourly timesteps for an exemplary scenario where initial internal (T_{ext1}) and external (T_{ext2}) wall surface temperature conditions (time $t = 0$) started at 10.73 and 10.82°C, respectively. This exemplary initial state scenario could correspond to a situation where an indoor zone experiences heat loss after being recently ventilated (e.g., due to open window for long period of time) causing the internal zone air temperature to drop closer toward the outdoor temperature.

The behavior of the trajectory observed in Figure 3b can be described as a linear combination of thermal conduction dynamics between T_{ext1} and T_{ext2}, and the influence of the ambient temperature T_{ext}, as described by the state equation in equation (7). In terms of control systems theory, T_{ext} can be described as a control forcing signal. Therefore, if we decouple the T_{ext} from T_{ext1} and T_{ext2}, we reveal the trajectory traced by the conduction dynamics alone (blue curve in Figure 4a). We do this by fixing the ambient outdoor temperature to a constant value (0°C), that is, we consider $x(t+1) = Ax$.

The geometry of the decoupled trajectory describes how the full state of the system approaches steady-state conditions, that is, thermal equilibrium. We can interpret the geometry as follows: shortly after the room was ventilated (initial conditions at time $t = 0$), the external surface of the wall (T_{ext2}) experiences a sudden drop in temperature due to a decrease in outdoor ambient temperature. Note that how the internal surface temperature (T_{ext1}) starts to drop at a much slower rate due to the thermal resistance of the wall. Finally, the internal surface temperature drops toward zero very rapidly once the thermal capacity of the wall is reached. In other words, the path traced by the trajectory can be interpreted as the dynamical behavior of conduction given the specified thermophysical properties of the wall.

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

The velocity and path of the traced trajectory are governed by an underlying vector field which describes the global behavior of any trajectory in the state space. The vector field can be described fully by the governing dynamics matrix A in the SSM from equation (7). More specifically, the velocity and direction of any coordinate T_{ext1}, T_{ext2} in the state-space vector field is a function of the eigenvectors v_1, v_2 and eigenvalues λ_1, λ_2 composing A. In fact, any point in the state space $T = (T_{ext1}, T_{ext2})$ (representing a thermal scenario) can be described as a linear combination of the eigenvectors. The velocity in a thermal state space represents the rate of change of the temperature states over time and is dictated by the eigenvalues also interpreted as "thermal eigenvalues" (Madsen, 1995; Hyun and Wang, 2019). Thus, by generating solution trajectories for arbitrary initial state combinations $T(t_0) = [T_{ext1}(t_0), T_{ext2}(t_0)]$, we reveal the structural geometry of the dynamics governing the thermal behavior of the wall in Figure 4b (also known as a phase portrait). It is this innate initial condition generalizability that we aim to leverage for the KT (and this level of physical interpretability for the wider research goal).

Note that phase portraits are more typically used to plot the state space of continuous-time systems; however, the plots in this paper are for discrete-time systems as a result of the post matrix exponential of the dynamics matrix A in equation (12).

Figure 5a,b illustrate the state trajectories for heat transfer through thicker walls, of 0.6 and 1.5 m, respectively. On qualitative comparison with Figure 3a, we can instantly note a change in the geometry of the vector field and consequently the trajectories, governed by a rotational transformation in the respective orthogonal eigenvector pairs. As the wall gets thicker and the combined U-value decreases, the gradient of heat transfer decreases. When comparing the state-space geometry for different wall systems, for example, varying thicknesses, we can instantly observe a mappable/interpolative similarity. The geometric similarity of these independent but related systems begs the question if mechanistic knowledge can be transferred across a range of thermal systems if represented geometrically as such. Furthermore, Figures 15 to 17 in the Supplementary Material illustrate the geometric differences between dynamics of a system derived from first principles and dynamics derived from data dynamics using spectral decomposition techniques. This view opens up a geometric way to approach/develop geometry-based calibraiton techniques. This is the core motivation behind our work and why we look toward DA as a means to transfer mechanistic knowledge.

4. Physics-Based Domain Adaptation

4.1. Source and target data pre-processing (step 2)

We start by generating a source dataset $X_s \in \mathbb{R}^{n \times z}$, using the physics-based SSM which is computationally inexpensive to run. On the other hand, we assume availability of a set of measurement data $X_t \in \mathbb{R}^{n \times z}$

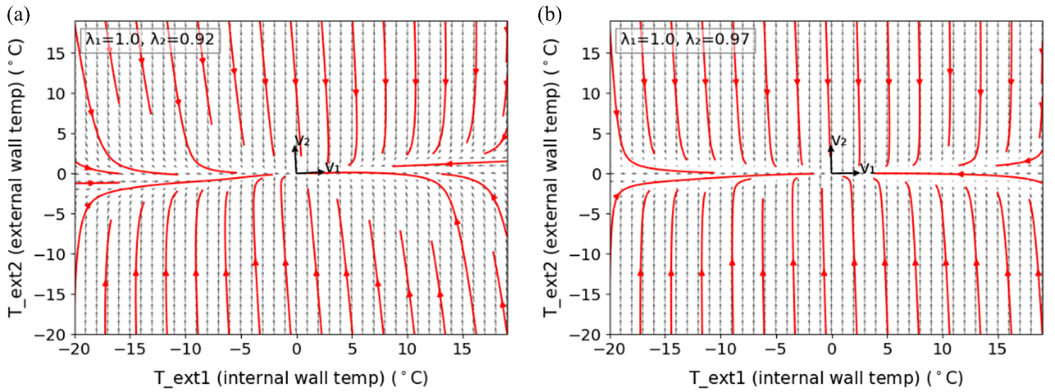

Figure 5. *(a) Heat transfer through 0.6 m thick wall. (b) Heat transfer through 1.5 m thick wall.*

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

which, for this paper, we generate synthetically (from an EnergyPlus model of the same system) as a representation of empirically observed building measurement data. It is our goal to learn a matrix that geometrically aligns X_s with X_t. Note that in this approach, we assume pairwise correspondence between source and target data points where both are "generated" under the same input conditions.

Subsequently, X_s and X_t are split according to a training/testing ratio (l/m). The target training set is used to derive the target subspace eigenvectors V_t via one of the ROM methods discussed, and subsequently also used to learn the transformation matrix M in step 4. On the other hand, the target test set is used to validate the forecasted data. The size of V_t used to train the ROM may be set independently from the size of V_t used to learn the transformation.

Prior to subspace embedding, the source and target data are centered along the zero mean as specified by equation (14). Centering the data is an important step and critical for SA as it brings closer the source and target distributions before alignment. The training set and testing set are centered independently to avoid prediction bias due to data leakage from training information to testing.

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j = 0. \tag{14}$$

4.2. Infer source subspace (step 3)

In contrast to standard SDA approaches, we derive the basis of the source subspace as eigenvectors V_s inferred directly from the physics-derived low-rank SSM representation of the energy system in step 1, via eigendecomposition of its system matrix A. The decomposition is described in terms of the eigenvalue problem in equation (15) where λ_s is a diagonal matrix containing the eigenvalues. In typical SDA applications, the eigenvectors are derived directly from source data via ROM methods and represent the principal directions in which the variance spans. In our case, the eigenvectors (and eigenvalues) hold direct connotation with the governing structure of the state space as illustrated in Section 2 and thus, aid to improve generalizability of the learned transformation when embedding the measurement data into (lower dimensions) subspaces. It is worthy to note that the A derived from the thermal RC network is symmetric implying that the eigenvectors are linearly independent and orthogonal.

$$A = V_s \lambda_s V_s^{-1}. \tag{15}$$

4.3. Infer target subspace (step 4)

In our approach, we derive the vector basis of the target subspace V_t, directly from the available measurement data X_t via ROM. In this paper, we adopt POD as the ROM method of choice.

POD is a linear and unsupervised ROM method widely employed in various fields including signal analysis and pattern recognition. POD is equivalent to the well-known PCA in that they both employ SVD to decompose the matrix into orthogonal eigenvectors, aka orthogonal modes. These modes are then used to project the original dataset into an optimal lower-dimensional[1] approximation. When applied to fields pertaining to dynamic behavior, the eigenvectors obtained are interpreted as characteristic spatial structure also known as modes, which refer to the characteristic structure of the source system, especially in flow dynamics.

In typical SA applications, further method extensions are implemented to account for deriving the ideal dimensionality of source and target subspaces when applying ROM to the source and target data. The same does not apply to our proposed approach, where both source and target eigenvectors correspond directly to the states of the system thus, already exist in the governing dimensionality.

[1] While in our demonstrative application in Section 5, we deal with a simple system that does not require order reduction, our approach applies with reduced-order states.

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

4.4. Subspace alignment (step 5)

The focus of most SDA methods is to learn an optimal matrix transformation operator/s that aligns the source and target subspaces. We explore two alternative sub-methods for SA: standard SA via minimization of the Bergman divergence perspective (Fernando et al., 2013) and Procrustes-based SA (Wang and Mahadevan, 2008). A summary of both algorithms used is given by Algorithms 1 and 2 and are further discussed in more detail here.

The classic SA method is the standard DA alignment approach where a linear transformation that maps source subspace to the target one is determined. This is achieved by aligning the basis vectors using a transformation matrix M from X_s to X_t ($M \in \mathbb{R}^{d\times d}$) which is learned by minimizing the Bergman matrix divergence in equations (16) and (17), where $\|.\|_F^2$ is the Frobenius norm. The Frobenius norm is invariant to orthonormal operations and thus can be rewritten as the objective function in equation (19) as suggested by Fernando et al. (2013) and Sebban et al. (2014).

$$F(M) = \|X_s M - X_t\|_F^2, \tag{16}$$

$$M^* = argmin_M F(M), \tag{17}$$

$$M^* = {X_s}^T X_t, \tag{18}$$

$$F(M) = \|X_s M - X_t\|_F^2. \tag{19}$$

In the second alignment approach, we adopt Procrustes analysis, which is a classic statistical method typically used for shape analysis and image registration of 2D/3D data, which seeks isotropic dilation and the rigid translation, reflection, and rotation needed to best match one data structure to another. More specifically, Procrustes alignment removes the translational, rotational, and scaling components so that the optimal alignment between source–target instance pairs is found. Instead of finding a single transformation matrix M, Procrustes manifold alignment (Wang and Mahadevan, 2008; Perron et al., 2021) seeks to align source and target subspaces using the rotation matrix $r \in \mathbb{R}^{n\times n}$and scaling factor s, found via Procrustes analysis on the embedded data. First, we project SSM-generated data onto the source eigenvectors, and the target measurement data onto the target eigenvectors,

$$\tilde{X}_s = X_s {V_s}^T, \quad \tilde{X}_t = X_t {V_t}^T. \tag{20}$$

Subsequently, we align the two using ordinary Procrustes via the SVD of the embedded datasets:

$$U\Sigma V^T = SVD\left({\tilde{X}_t}^T \tilde{X}_s\right), \tag{21}$$

$$r = UV^T, \tag{22}$$

$$s = \frac{trace(\Sigma)}{trace\left(\tilde{X}_s \left(\tilde{X}_s\right)^T\right)}, \tag{23}$$

where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{n\times n}$ are matrices containing the left and right singular vectors and $\Sigma \in \mathbb{R}^{n\times 1}$ is a diagonal matrix containing the singular values on its diagonal. Note that in this implementation, we ignore the translation term during Procrustes since given that both $\tilde{X}_s$and $\tilde{X}_t$ are centered to zero mean prior to subspace embedding. Finally, we reconstruct the original training data by applying r and s to the embedded source data (equation (24)) and lifting it into the original target space (equation (25)). With this in place, we can forecast for new inputs beyond the available target data.

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

$$\tilde{X}_a = sr\tilde{X}_{s,} \tag{24}$$

$$NX_a = V_s\tilde{X}_a{}^T. \tag{25}$$

Algorithm 1. SA-based DA via Bergman divergence.

Inputs: SSM-generated data X_s, measurement data (simulated) X_t.
Output: Forecasted target data $X_{\hat{t}}$.

1. Infer source eigenvectors V_s from the SSM by decomposing A, where $A = V_s\lambda_s V_s{}^{-1}$.
2. Infer target eigenvectors V_t from measurement data via POD.
3. Obtain optimal M by aligning V_s toward V_t via minimization of the Bergman divergence, where $M = argmin_M\left(\|V_sM - V_t\|_F^2\right)$.
4. Apply M to obtain the basis of the new target-aligned subspace V_a, where $V_a = V_s\mathrm{M}$.
5. Project source data X_s onto new target-aligned subspace V_a, where $\tilde{X}_a = X_sV_a{}^T$.
6. Reconstruct aligned source data in target domain, where $X_a = \tilde{X}_aV_t$.
7. Forecast $X_{\hat{t}}$ for new inputs by repeating steps 3, 6, and 7.

Algorithm 2. SA-based DA via Procrustes analysis.

Inputs: SSM-generated data X_s, measurement data (simulated) X_t.
Output: Forecasted target data $X_{\hat{t}}r$.

1. Infer source eigenvectors V_s via decomposition of state transition matrix A into $V_s\lambda_s V_s{}^{-1}$.
2. Infer target eigenvectors V_t from data via POD.
3. Project source data X_s onto source subspace V_s, where $\tilde{X}_s = X_sV_s{}^T$.
4. Project target data X_t onto target subspace V_t, where $\tilde{X}_t = X_tV_t{}^T$.
5. Infer rotation matrix r, scaling factor s, and translation factor t via Procrustes analysis, where $r,s,t = argmin\left(\|X_s - kX_tr\|_F\right)$.
6. Apply r,s,t transformations to $\tilde{X}_s$ via $\tilde{X}_a = rs\tilde{X}_s + t$.
7. Reconstruct aligned target data via $X_a = V_t\tilde{X}_a{}^T$.
8. Forecast $X_{\hat{t}}$ for new inputs by repeating steps 3, 6, and 7.

4.5. Error metrics

We quantify the performance of the SA-based DA by comparing its prediction accuracy for out of sample inputs, to that of the target data which we consider to be the "true" observed data. More specifically, we compute the coefficient of variation of root-mean-squared error (RMSE) and the normalized mean bias error (NMBE) as follows:

$$CV(RMSE)(X) = \frac{1}{\overline{X}}\sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\left(X_i - \widehat{X}_i\right)^2}, \tag{26}$$

$$NMBE(X) = \frac{\sum_{i=0}^{n-1}\left(\widehat{X}_i - X_i\right)}{\sum_{i=0}^{n-1}(X_i)}. \tag{27}$$

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

We adopt the guidelines set by the building energy modeling and forecasting community (ASHRAE, 2014) to determine the validity of the RMSE and NMBE results for energy forecasting. For hourly model calibration, it is recommended that maximum allowed NMBE is capped at 10% while the CV(RMSE) at 30%.

5. Demonstrative Application

As aforementioned, we illustrate two application scenarios with the demonstrative wall example: (a) calibration of physics-simulated data to measurement data, and (b) reuse of physics-based model for a variation of wall types undergoing similar phenomenon. For each scenario, we plot and report pre- and post-alignment results via CV(RMSE) and NMBE measures for timeseries reconstruction and forecasting.

5.1. Transfer across equivalent systems (calibration)

We first simulate 7,000 hr of synthetic measurement temperature data for a 0.2 m wall scenario using a high-fidelity simulation model of the thermal zone in Figure 2 using EnergyPlus, which is a widely simulation software used by the building energy community for thermal modeling and simulation. The goal is to calibrate data simulated using a physics-derived SSM model of the same wall scenario (source), closer toward the simulated measurement data (target).

For this study, we select varying training sets from the simulated high-fidelity data: 2,000, 4,000, 6,000 hr. During the simulation, we record the temperature on the inner surface T_{ext1} and exterior surface T_{ext2} of the wall at an hourly timestep. Our goal in this exercise is to forecast beyond the measurement data assigned as training ($T_{ext1}target$, $T_{ext2}target$) by leveraging response data ($T_{ext1}source$, $T_{ext2}source$) generated from a governing physics-based description (SSM) of the same 0.2 m thick wall scenario, and which is computationally inexpensive to obtain. Note that for clarity, throughout this study we illustrate only the alignment and forecast for the temperature of the inner surface of the wall T_{ext1} since it is significantly influenced by the conductive dynamics through the wall and thus, of main interest.

We first specify the elements of the system matrix A by substituting the thermophysical values for a wall of 0.2 m thickness in Table 1 into equation (9). Subsequently, we obtain the state transition matrix Φ_A for an hourly timestep dt by solving e^{dtA} (equation (29)). Here, we assume an hourly timestep $dt = 3{,}600s$. Given Φ_A, we can obtain the eigenvectors V_{s1}, V_{s2} (specified in Table 5 in the Supplementary Material), via the decomposition in equation (12).

$$A = \begin{bmatrix} -1.2019\text{e}-05 & 1.2019e-05 \\ 1.2019e-05 & -7.879e-05 \end{bmatrix}, \tag{28}$$

$$\Phi_A = e^{dtA} = \begin{bmatrix} 0.95848 & -0.03684 \\ -0.03684 & 0.75379 \end{bmatrix}. \tag{29}$$

Once we compute the basis eigenvectors of the source subspace, we proceed to extract the basis eigenvectors V_{t1}, V_{t1} of the target subspace from the available measurement data using data-driven ROM. Here, we use POD to infer the basis of the target subspace. The target eigenvectors derived via POD are specified in Table 5 in the Supplementary Material.

Now that we have obtained the vector bases for both source and target subspaces, we first project the data generated by the SSM ($T_{ext1}source$, $T_{ext2}source$) onto the source subspace (onto V_{s1}, V_{s2}), then project the measurement data ($T_{ext1}target$, $T_{ext2}target$) onto the target subspace (onto V_{t1}, V_{t2}), and subsequently proceed to learn a transformation map by aligning the projected source data closer toward the projected target data using the (a) standard SA via Bergman divergence minimization and (b) Procrustes-based SA, described earlier. To reiterate, with (a), we learn a rotational matrix only, while with (b), we learn a rotational matrix, together with scaling and translation factors. Once obtained, the

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

Table 2. *CV(RMS) accuracy for forecasting 1,000 hr using standard SA via Bergmann divergence.*

	Standard SA (Bergmann divergence)	
Training size (hr)	0	
Thickness (meters)	$0.2_{ssm} \rightarrow 0.2_{true}$	$0.8_{ssm} \rightarrow 0.8_{true}$
CVRMS SSM (°*C*)	8.74%	6.34%
$CVRMS_{POD}$ aligned (°*C*)	23.99%	10.66%

transformation map is first applied to reconstruct the target measurement data used for training by aligning the source data to the target data, and subsequently used to geometrically transform new data generated by the physics-derived SSM into target-aligned data to forecast for 1,000 hr beyond the target measurement data. The CV(RMSE) forecasting results via the standard SA Bergman divergence and Procrustes SA approaches are tabulated in Tables 2 and 4, respectively, to aid comparison between the two alignment approaches for a variety of transfer scenarios. The standard SA Bergmann approach performs poorly at the get-go as can be noted by the CV(RMSE) results. For this reason, in this section, we discuss and illustrate plots for Procrustes-based alignment only.

Figure 6 illustrates the reconstruction of the internal wall surface temperature T_{ext1} via Procrustes-based alignment of the physics-derived T_{ext1} data (red line). Subsequently, the learned transformation was applied to forecast a further 1,000 hr (dotted red line). We note an overall post-alignment improvement in the forecasting CV(RMSE) and NMBE values. We observed that while the physics-derived SSM already approximates the measured system quite well with a CV(RMSE) and NMBE within a desirable bound ($\leq$30 and $\leq$10%, respectively), our SDA approach improved further these forecasting accuracy values by 1.56 and $-$3.57%. In the following figures, we also plot the reconstruction and forecasting error (bottom) as a function of localized difference between the aligned and target datapoints (error_postaligned). We compare this with the localized difference between the output from the physics-derived SSM and the target datapoints (error_prealigned).

We repeated the calibration study for a thicker wall (0.8 m) using Procrustes-based SA on 2,000 hr training data. From Table 3, we note that the forecasting accuracy of the aligned (transformed) source data in the 0.8 m wall calibration scenario is 4.49% worse than the original physics-derived SSM (6.34%). We observed improvement for this case when increasing the training data size.

5.2. Transfer across different but related systems (reuse)

We hypothesize that building energy systems governed by similar phenomena have geometrically similar state-space structures implying that their governing subspaces live in a mutual latent space. This suggests that mechanistic descriptions of energy transfer through one wall could be used to describe and forecast the energy transfer through that of a wall with different geometric and material properties, undergoing similar energy transfer phenomena. In this context, we demonstrate how the SDA approach can be used to leverage the low-rank physics-derived SSM of one wall to predict and forecast the internal wall surface temperatures of observed walls with varying thickness and material properties, by learning a SA transformation between physics and given wall surface temperature measurement data.

Figure 7 illustrates the adaptation of a physics-derived SSM for a 0.6 m wall to forecast beyond data observed from a 0.2 m wall via the POD-derived target subspace. The respective source and target eigenvectors can be found in Table 6 in the Supplementary Material.We can immediately observe how the transformation learned between the physics-derived data and the target data has successfully been captured by its ability to reconstruct the training set (red line) when applied to the source data and subsequently to forecast a further 1,000 hr (dotted red line). We utilize the CV(RMSE) and NMBE values to compare the prediction improvement while ensure we are still in desirable bounds typically adopted for building energy model calibration ($\leq$30 and $\leq$10%, respectively). The improvement can also be noted by observing the difference in pre- and post-alignment error timeseries in the bottom plots.

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

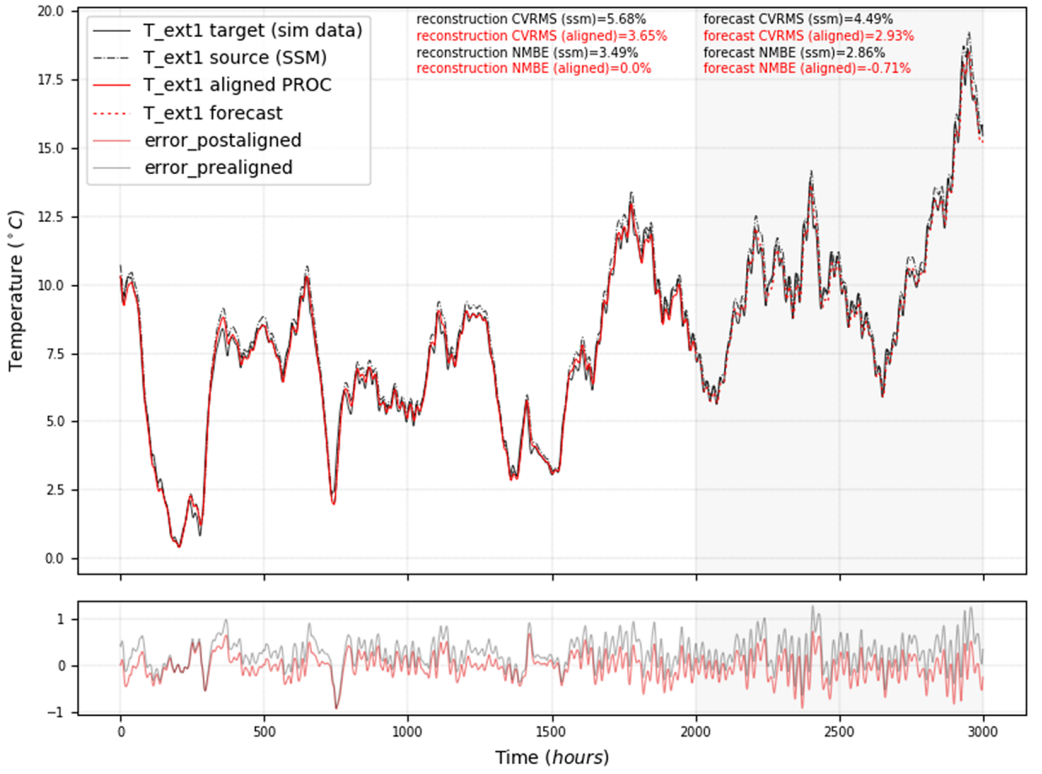

Figure 6. *Calibration scenario: 0.2 m wall SSM (source) to 0.2 m wall data (target). The target subspace was derived via POD (orthogonal eigenvectors).*

Table 3. *CV(RMS) accuracy for forecasting 1,000 hr using Procrustes-based SA for varying training size (2,000 hr, 4,000 hr, 6,000 hr).*

	Procrustes-based SA		Procrustes-based SA		Procrustes-based SA	
Training size (hr)	2,000		4,000		6,000	
Thickness (meters)	$0.2_{ssm} \rightarrow 0.2_{true}$	$0.8_{ssm} \rightarrow 0.8_{true}$	$0.2_{ssm} \rightarrow 0.2_{true}$	$0.8_{ssm} \rightarrow 0.8_{true}$	$0.2_{ssm} \rightarrow 0.2_{true}$	$0.8_{ssm} \rightarrow 0.8_{true}$
CVRMS SSM	4.49%	7.53%	2.75%	3.57%	2.45%	12.00%
$CVRMS_{POD}$ aligned	2.93%	5.05%	2.25%	6.07%	1.76%	9.81%

On the other hand, Figure 8 illustrates the adaptation of a physics-derived SSM for a 0.2 m wall to forecast beyond data observed from a 0.6 m wall. While we note forecasting improvement indicating a closer match to the target measurement data, we observe that the transformation via Procrustes alignment does not dampen the noisier source signal. In Section 5.3, we take a closer look to interpret why this occurs.

Lastly, Figure 9 illustrates the adaptation of the SSM for a red brick wall of 0.8 m thickness, 0.72 W/mK conductivity, 1,920kg/m^3 material density, and 780 J/kgK specific heat capacity, to forecast

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

Table 4. *CV(RMS) accuracy for forecasting 1,000 hr using cross-domain Procrustes-based SA for varying wall thicknesses only.*

	Procrustes-based SA					
Training size (hr)	2,000					
Thickness (meters)	$0.2_{ssm} \rightarrow 0.6_{true}$	$0.2_{ssm} \rightarrow 0.8_{true}$	$0.2_{ssm} \rightarrow 0.9_{true}$	$0.6_{ssm} \rightarrow 0.2_{true}$	$0.8_{ssm} \rightarrow 0.2_{true}$	$0.9_{ssm} \rightarrow 0.2_{true}$
CVRMS SSM	18.45%	23.02%	23.59%	19.88%	23.5%	25.09%
$CVRMS_{POD}$ aligned	13.38%	16.75%	16.82%	5.63%	5.99%	11.85%

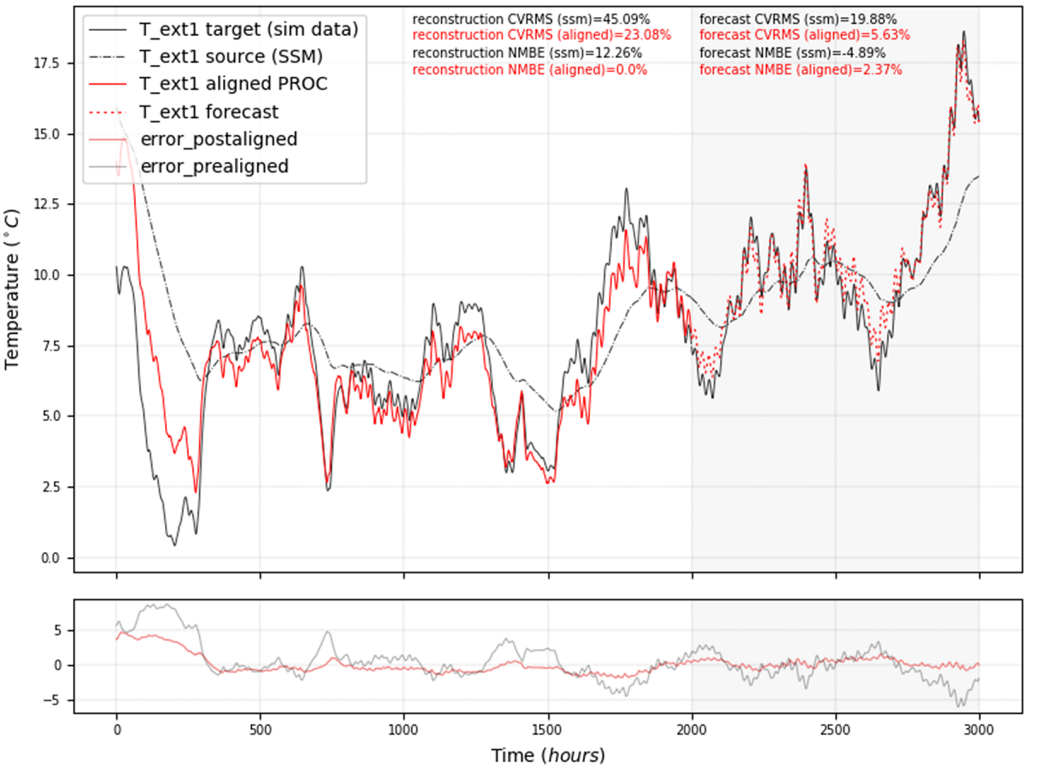

Figure 7. *Cross-domain generalization scenario: 0.6 m wall SSM (source) to 0.2 m wall data (target). The target subspace was derived via POD (orthogonal eigenvectors).*

beyond temperature data measured from a concrete wall of 0.8 m thickness, 1.3 W/mK conductivity, 2,240kg/m^3 material density, and 840 J/kgK specific heat capacity. The respective source and target eigenvectors can be found in Table 7 in the Supplementary Material. We can immediately observe that the transformation learned between the SSM and target data is able to forecast the internal surface temperature of the concrete wall for unseen external temperature inputs. This can be noted by the improvements on CV(RMSE) and NMBE values.

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

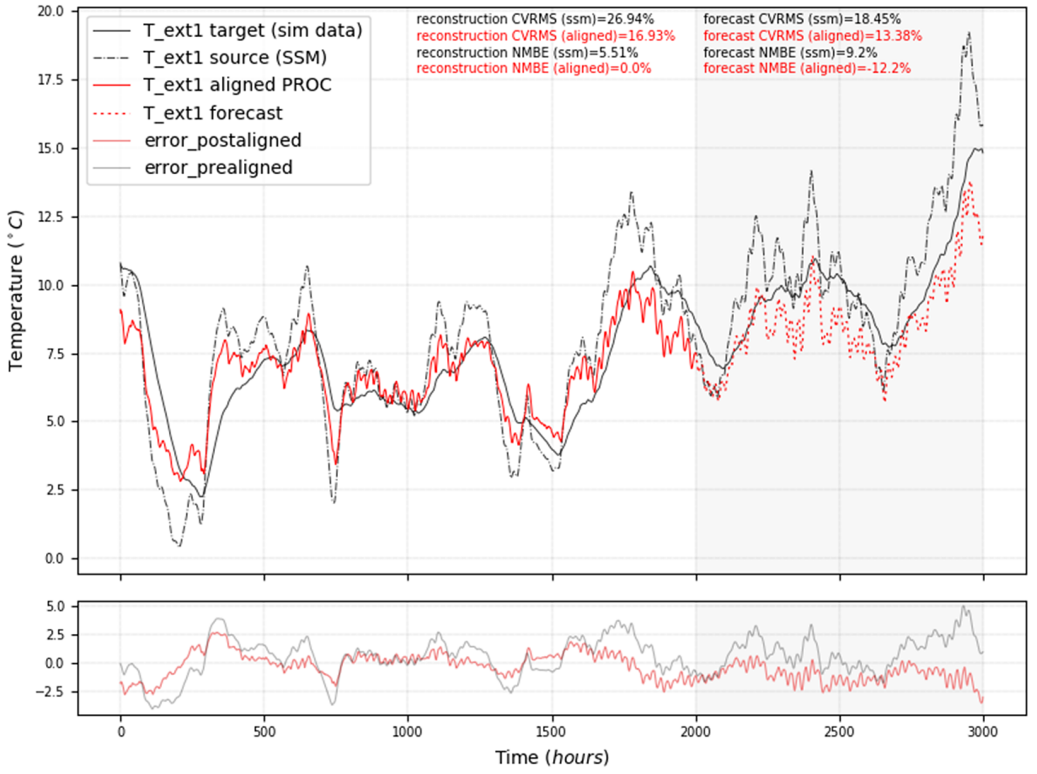

Figure 8. *Cross-domain generalization scenario: 0.2 m wall SSM (source) to 0.6 m wall data (target). The target subspace was derived via POD (orthogonal eigenvectors).*

5.3. Discussion

The standard SA method via Bergmann divergence minimization does not depend on data because it acts only on the subspaces while Procrustes-based SA acts directly on the local geometry of the projected source and target data. The former only accounts for rotational transformational between source and target subspaces while the latter yields a rotation matrix, skewing, translation, and reflection factors. In Section 5.1 we observed that rotation only is not sufficient to completely describe geometric mappings between subspace-embedded source and target data, particularly in cases where the source and target eigenvectors are in a mirrored orientation from each other. Thus, while we acknowledge that Procrustes-based SA is data-dependent, it is suitable for capturing geometric transformations between physics-derived and measurement timeseries signals, even when transferring across different but related wall-types (Table 4).

We observed that when lifting the target-aligned data from the embedded space to the target space, the source–target alignment can be compromised in cases when the basis of the target-aligned subspace does not align well with the basis of the target subspace. For example, in the cross-domain alignment example ($0.2_{ssm} \rightarrow 0.9_{true}$) illustrated in Figure 10, top, we instantly note that Procrustes successfully aligns the source with target signal in the embedded space (left) but the alignment is subsequently compromised when lifting into the target subspace (right). This was evident when comparing the source–target alignment data in the embedded space. Figure 10, bottom, illustrates another instance when this occurs, this time a same-domain alignment scenario.

We also experimented with dynamic mode decomposition (DMD) as an alternative ROM method to infer the target subspace. DMD is known to be superior at capturing the governing temporal structure underlying data. DMD was not suitable for all domain transfer cases mainly since the eigenvectors inferred via DMD are non-orthogonal. It is noteworthy that in the successful cases, we observed

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

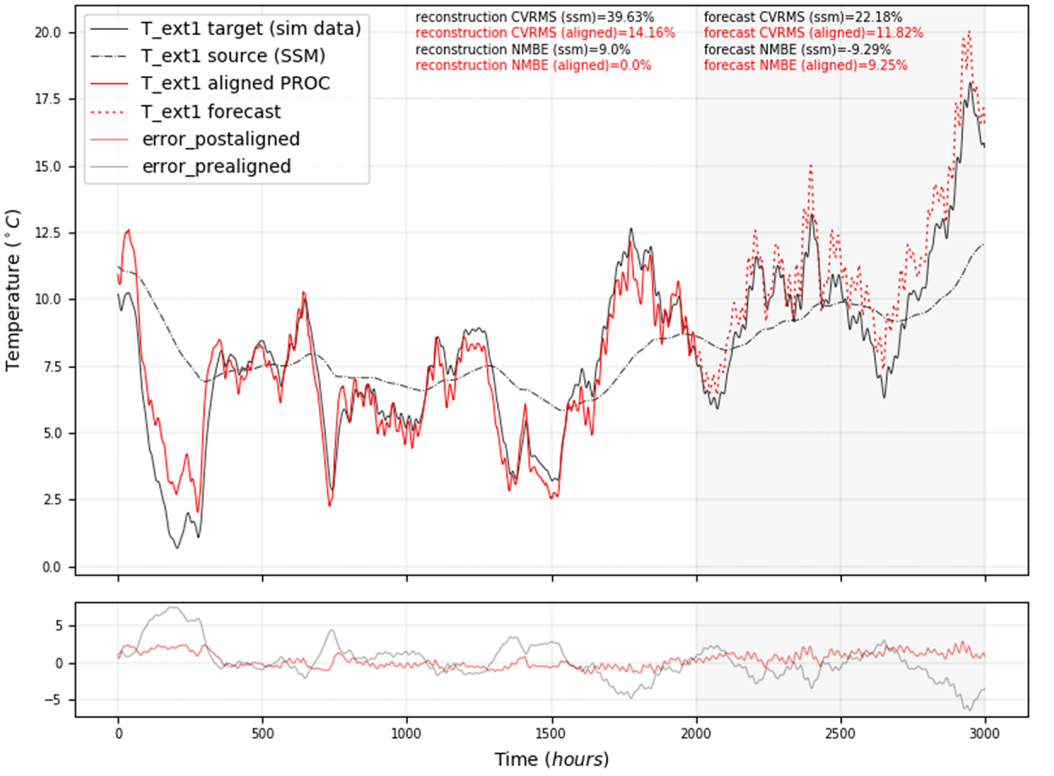

Figure 9. *Cross-domain generalization scenario: 0.8 m red brick wall SSM (source) to 0.3 m concrete wall data (target). The target subspace was derived via POD (non-orthogonal eigenvectors).*

significant alignment improvement resulting in longer-term forecasts when compared to POD-derived target subspace. This can be observed by comparing Figure 11 in the Supplementary Material with Figure 12 in the Supplementary Material.

While the demonstrative example in this paper assumes a simple building component, the physics-based SDA framework is scalable to model more intricate building systems and or thermal zones. There exists dedicated work to automate the generation of RC networks from 3D models of thermal zones (Kim et al., 2018). Furthermore, the energy balance equations assumed to inform the SSM are well known by the energy community and straightforward to implement.

While the examples illustrated above do not assume noisy conditions in the measurement data, in Figures 13 and 14 in the Supplementary Material, we illustrate a scenario where random noise was added to the measurement data. In noisy measurement conditions, the modeling goal is not to forecast a noisy signal but rather, to forecast a denoised signal. Thus, bypassing a noisy measurement signal wholly depends on the capability of the ROM method to decouple the noise from the true signal. In the example illustrated by Figures 13 and 14 in the Supplementary Material, we introduce noise sampled from a random distribution with a mean of 0.5 and a standard deviation of 0.9, to the measurement data from the scenario discussed in Section 5.2. We can observe how the SDA approach is able to learn a transformation between the source and a noise-filtered target signal that generalizes successfully beyond measured timesteps. This is largely due to the SVD decomposition of the noisy signal when applying POD to the target data where the SVD acts as a noise filter (Shin et al., 1999; Epps and Krivitzky, 2019). Figures 13 and 14 in the Supplementary Material demonstrate the SA results in the subspace-embedded and the lifted space, respectively. If POD does not suffice for denoising the timeseries data, there exist extensions of POD such as SPOD (Sieber et al., 2016) that focus on spectral decomposition to deal with such

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

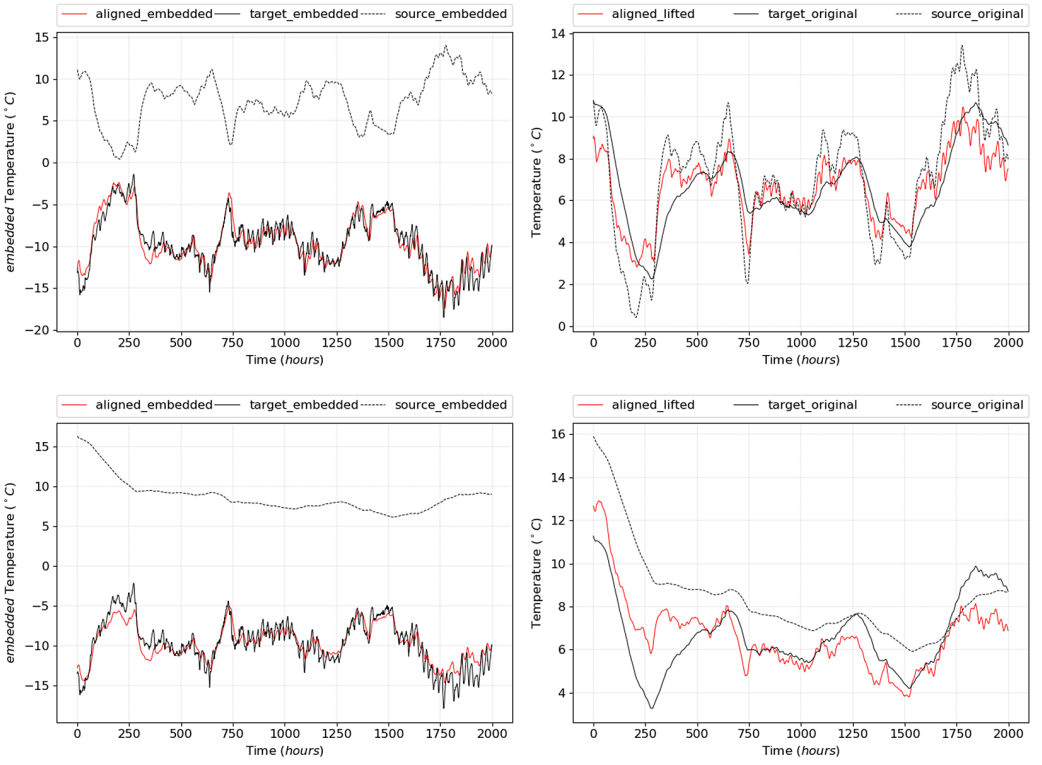

Figure 10. *Top: cross-domain alignment: 0.2 m wall SSM (source) to 0.9 m wall data (target). Middle: same-domain alignment: 0.8 m wall SSM (source) to 0.8 m wall data (target). Left plots illustrate alignment in the embedded space. Right plots illustrate the alignment after lifting from embedded space.*

decoupling/filtering more successfully. The main difference between POD and SPOD is that the modes vary in both space and time and are orthogonal under space–time inner product, rather than only spatially.

For stochastic LTI SSMs, the presented SDA approach could be extended by introducing a stochastic noise term to a continuous-time LTI SSM which can subsequently be translated in to discrete-time via integration (Madsen, 1995). Such an implementation would not alter the state transition matrix (A matrix) from which we derive the source subspace eigenvectors.

The "adaptability" of a physics-based model to forecast for a target measured system using our presented SDA approach is limited by two factors: (a) The state dimensions in the physics-based model need to be equivalent to the dimensions of the measurement data implying that the resolution of the SSM is limited by the number of available corresponding measurements. In future work, we intend to facilitate prediction of unobserved target measurements by leveraging their corresponding physics-derived state in the SSM. (b) We assume source and target data have pairwise correspondence, implying that the SSM control inputs are the same (measured) control signals known to be influencing the target measurement data, for example, the external ambient temperature from the case study in Section 5. Pairwise correspondence can easily be ensured for MPC applications because most typical building energy scenarios can be described by well-established governing laws, even if approximate.

6. Conclusion

In this paper, we introduced a novel TL approach to address better model-generalization for timeseries forecasting of building energy, by exploiting well-established ODEs governing thermal transfer in a data-driven framework. Our approach combines physics-derived LTI SSMs with SA where a geometric

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

transformation is sought to align the subspace derived from governing physics closer toward the subspace decomposed from the measurement data. Subsequently, the transformation is used to forecast beyond the observed measurement data by transforming physics simulated data for unseen timesteps into approximated measurement data.

While our physics-based SDA framework addresses generalization for unseen timesteps (timeseries forecasting), it can easily be adopted for cross-system generalization where, for example, we can exploit LTI SSMs identified from data (instead of first principles) of a fully observed system to forecast for unseen timesteps given measurement data from a different but related system. Such an approach could benefit data-efficient control strategies across thermal zones or building components within the same building, by transforming LTI SSM models calibrated on specific thermal zones/components to forecast for other different but related thermal zones/components.

In subsequent work, we intend to test the physics-based SDA framework on real observed scenarios and applications pertaining to energy systems and components in buildings. This work stems from a broader research goal where we aim to overcome persisting caveats with traditional data-driven energy modeling and forecasting, namely, generalizability, data-dependency, and interpretability.

Acknowledgments. We are grateful for the technical assistance of Dr. Andrew Duncan.

Author contribution. Conceptualization: Z.X.C., R.C.; Data visualization: Z.X.C.; Formal analysis: Z.X.C., L.M.; Methodology: Z.X.C.; Software: Z.X.C.; Validation: all authors; Writing original draft: Z.X.C. All authors approved the final submitted draft.

Competing interest. The authors declare no competing interests exist.

Data availability statement. The data used throughout this paper were synthetically generated. The authors have ensured to provide sufficient information to replicate the analytical and numerical models used to generate the synthetic data.

Funding statement. This work was supported by The Alan Turing Institute's Data-Centric Engineering Program under the Lloyd's Register Foundation Grant G0095, and Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1 and EPSRC Grant EP/W006022/1, particularly the digital twins in engineering theme within those grants and The Alan Turing Institute.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/dce.2023.8.

References

Amara F, **Agbossou K**, **Cardenas A**, **Dubé Y and Kelouwani S** (2015) Comparison and simulation of building thermal models for effective energy management. *Smart Grid and Renewable Energy 6*(4), 95–112. https://doi.org/10.4236/sgre.2015.64009

ASHRAE (2009) ASHRAE Handbook - Fundamentals (I-P edition) - Energy estimating and modeling methods.

ASHRAE (2014) ASHRAE Guideline 14–2014: Measurement of Energy, Demand, and Water Savings. Available at www.ashrae.org (accessed 13 April 2022).

Baktashmotlagh M, **Harandi MT**, **Lovell BC and Salzmann M** (2013) Unsupervised domain adaptation by domain invariant projection. In *Proceedings of the IEEE International Conference on Computer Vision*, Sydney, NSW, Australia: IEEE. https://doi.org/10.1109/ICCV.2013.100

Blitzer J, **Mcdonald R and Pereira F** (2006) Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Sydney, Australia: Association for Computational Linguistics, Sydney, Australia.

Bourdeau M, **Zhai Xq**, **Nefzaoui E**, **Guo X and Chatellier P** (2019) Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society 48*, 101533. https://doi.org/10.1016/j.scs.2019.101533

Bueno B, **Norford L**, **Pigeon G and Britter R** (2012) A resistance-capacitance network model for the analysis of the interactions between the energy performance of buildings and the urban climate. *Building and Environment 54*, 116–125. https://doi.org/10.1016/j.buildenv.2012.01.023

Candanedo JA, **Dehkordi Vahid R and Lopez P** (2013) A Control-Oriented Simplified Building Modelling Strategy. Available at http://task40.iea-shc.org/data/sites/1/publications/T40A52-Control-Oriented-Building-Modelling-Strategey-1057.pdf (accessed 02 February 2022).

Elhadji-Ille-Gado N, **Grall-Maes E and Kharouf M** (2017) Transfer learning for large scale data using subspace alignment. In *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, 2017-December*. Cancun, Mexico: IEEE. pp. 1006–1010. https://doi.org/10.1109/ICMLA.2017.00-20

Epps BP and Krivitzky EM (2019) Singular value decomposition of noisy data: Noise filtering. *Experiments in Fluids 60*(8), 126. https://doi.org/10.1007/s00348-019-2768-4

Fang X, **Gong G**, **Li G**, **Chun L**, **Li W and Peng P** (2021) A hybrid deep transfer learning strategy for short term cross-building energy prediction. *Energy 215*, 119208. https://doi.org/10.1016/j.energy.2020.119208

Fateh A, **Borelli D**, **Spoladore A and Devia F** (2019) A state-space analysis of a single zone building considering solar radiation, internal radiation, and PCM effects. *Applied Sciences 9*, 832. https://doi.org/10.3390/app9050832

Fayazbakhsh MA, **Bagheri F and Bahrami M** (2015) A resistance-capacitance model for real-time calculation of cooling load in HVAC-R systems. *Journal of Thermal Science and Engineering Applications 7*(4), 041008. https://doi.org/10.1115/1.4030640

Fernando B, **Habrard A**, **Sebban M**, **Tuytelaars T and Tuytelaars T** (2013) Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of ICCV*. Available at https://hal.archives-ouvertes.fr/hal-00869417 (accessed 07 March 2021).

Gao Y, **Ruan Y**, **Fang C and Yin S** (2020) Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. *Energy and Buildings 223*, 110156. https://doi.org/10.1016/J.ENBUILD.2020.110156

Goyal S and Barooah P (2011) A method for model-reduction of nonlinear building thermal dynamics. In *Proceedings of the 2011 American Control Conference*, San Francisco, CA, USA: IEEE.

Goyal S and Barooah P (2012) A method for model-reduction of non-linear thermal dynamics of multi-zone buildings. *Energy and Buildings 47*, 332–340. https://doi.org/10.1016/j.enbuild.2011.12.005

Goyal S, **Liao C and Barooah P** (2011) Identification of multi-zone building thermal interaction model from data. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, FL, USA: IEEE.

Haves P, **Norford LK and DeSimone M** (1998) ASHRAE 825-rp final report: A standard simulation test bed for the evaluation of control algorithms and strategies. Available at https://www.osti.gov/biblio/653219.

Hendricks E, **Jannerup O and Sorensen P** (2008) Linear systems control. In *Deterministic and Stochastic Methods*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-78486-9

Huang CH, **Yeh YR and Wang YCF** (2012) Recognizing actions across cameras by exploring the correlated subspace. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7583 LNCS (PART 1)*, Springer, Berlin, Heidelberg. pp. 342–351. https://doi.org/10.1007/978-3-642-33863-2_34

Hyun J and Wang S (2019) Systematically engineered thermal metastructure for rapid heat dissipation/diffusion by considering the thermal eigenvalue. *Applied Thermal Engineering 157*, 113487. https://doi.org/10.1016/J.APPLTHERMALENG.2019.03.058

Karpatne A, **Atluri G**, **Faghmous JH**, **Steinbach M**, **Banerjee A**, **Ganguly A**, **Shekhar S**, **Samatova N and Kumar V** (2017) Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering 29*(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Kim D, **Bae Y**, **Braun JE**, **Travis Horton W**, **Travis W and Horton TW** (2018) A Tool for Generating Reduced-Order Models from Building Energy Simulation Input Files to Enable Optimal Design and Control Analysis. Available at https://engineering.purdue.edu/ (accessed 17 January 2023).

Koeln J, **Keating B**, **Alleyne A**, **Price C and Rasmussen BP** (2018) Multi-zone temperature modeling and control. In *Advances in Industrial Control* (Issue 9783319684611. Cham: Springer International Publishing, pp. 139–166). https://doi.org/10.1007/978-3-319-68462-8_6

Kouw WM and Loog M (2018) An introduction to domain adaptation and transfer learning. Available at http://arxiv.org/abs/1812.11806 (accessed 21 June 2021).

Li X and Wen J (2014a) Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews 37*, 517–537). https://doi.org/10.1016/j.rser.2014.05.056

Lin Y, **Middelkoop T and Barooah P** (2012) Identification of control-oriented thermal models of rooms in multi-room buildings. In *51st IEEE Conference Decision and Control (CDC)*, Hawaii.

Madsen H (1995) Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings 22*, 67–79.

Mocanu E, **Nguyen PH**, **Kling WL and Gibescu M** (2016) Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning. *Energy and Buildings 116*, 646–655. https://doi.org/10.1016/J.ENBUILD.2016.01.030

Pan SJ, **Tsang IW**, **Kwok JT and Yang Q** (2011) Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks 22*(2), 199–210. https://doi.org/10.1109/TNN.2010.2091281

Pan SJ and Yang Q (2010) A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering 22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Perron C, **Rajaram D and Mavris DN** (2021) Multi-fidelity non-intrusive reduced-order modelling based on manifold alignment. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. *477*, 20210495. https://doi.org/10.1098/rspa.2021.0495

Picard D, **Jorissen F and Helsen L** (2015) Methodology for obtaining linear state space building energy simulation models. In *Proceedings of the 11th International Modelica Conference, Versailles, France, September 21–23, 2015*, Belgium: Liu Electronic Press, Vol. *118*, pp. 51–58. https://doi.org/10.3384/ecp1511851

Rätz M, **Javadi AP**, **Baranski M**, **Finkbeiner K and Müller D** (2019) Automated data-driven modeling of building energy systems via machine learning algorithms. *Energy and Buildings 202*, 109384. https://doi.org/10.1016/J.ENBUILD.2019.109384.

https://doi.org/10.1017/dce.2023.8 Published online by Cambridge University Press

Ribeiro M, **Grolinger K**, **ElYamany HF**, **Higashino WA and Capretz MAM** (2018) Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. *Energy and Buildings 165*, 352–363. https://doi.org/10.1016/j.enbuild.2018.01.034.

Sarkar T, **Rakhlin A and Dahleh MA** (2021) Finite time LTI system identification. *Journal of Machine Learning Research 22*, 1–61. Available at http://jmlr.org/papers/v22/19-725.html.

Sebban M, **Fernando B**, **Habrard A and Tuytelaars T** (2014) Subspace Alignment for Domain Adaptation. Available at https://www.researchgate.net/publication/265852615 (accessed 09 June 2021).

Sehovac L, **Nesen C and Grolinger K** (2019) Forecasting building energy consumption with deep learning: A sequence to sequence approach. In *Proceedings - 2019 IEEE International Congress on Internet of Things, ICIOT 2019 - Part of the 2019 IEEE World Congress on Services*, Milan, Italy: IEEE, pp. 108–116. https://doi.org/10.1109/ICIOT.2019.00029

Shao M, **Wang X**, **Bu Z**, **Chen X and Wang Y** (2020) Prediction of energy consumption in hotel buildings via support vector machines. *Sustainable Cities and Society 57*, 102128. https://doi.org/10.1016/J.SCS.2020.102128

Shin K, **Hammond JK and White PR** (1999) Iterative SVD method for noise reduction of low-dimensional chaotic time series. *Mechanical Systems and Signal Processing 13*(1), 115–124. https://doi.org/10.1006/MSSP.1998.9999

Sieber M, **Paschereit CO and Oberleithner K** (2016). Spectral proper orthogonal decomposition. *Journal of Fluid Mechanics*, *792*, 798–828. https://doi.org/10.1017/jfm.2016.103

Somu N, **M R**, **G R and Ramamritham K** (2020) A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy 261*, 114131. https://doi.org/10.1016/J.APENERGY.2019.114131

Tardioli G, **Kerrigan R**, **Oates M**, **O'Donnell J and Finn D** (2015) Data driven approaches for prediction of building energy consumption at urban level. *Energy Procedia 78*, 3378–3383. https://doi.org/10.1016/J.EGYPRO.2015.11.754

Wang W, **Hong T**, **Xu X**, **Chen J**, **Liu Z and Xu N** (2019) Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm. *Applied Energy 248*, 217–230. https://doi.org/10.1016/J.APENERGY.2019.04.085

Wang C and Mahadevan S (2008) Manifold Alignment using Procrustes Analysis. Computer Science Department Faculty Publication Series, 64.

Wang Z, **Wang Y**, **Zeng R**, **Srinivasan RS and Ahrentzen S** (2018) Random forest based hourly building energy prediction. *Energy and Buildings 171*, 11–25. https://doi.org/10.1016/J.ENBUILD.2018.04.008

Willard J, **Jia X**, **Xu S**, **Steinbach M and Kumar V** (2022) Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys 55*(4). https://doi.org/10.1145/3514228.

Cite this article: Xuereb Conti Z, Choudhary R and Magri L (2023). A physics-based domain adaptation framework for modeling and forecasting building energy systems. *Data-Centric Engineering*, 4, e10. doi:10.1017/dce.2023.8