
Physics-constrained Deep Recurrent Neural Models of Building Thermal Dynamics

Ján Drgoňa, Aaron R. Tuor, Vikas Chandan, & Draguna L. Vrabie

Pacific Northwest National Laboratory
Richland, Washington, USA

{jan.drgona, aaron.tuor, vikas.chandan, draguna.vrabie}@pnnl.gov

Abstract

We develop physics-constrained and control-oriented predictive deep learning models for the thermal dynamics of a real-world commercial office building. The proposed method is based on the systematic encoding of physics-based prior knowledge into a structured recurrent neural architecture. Specifically, our model mimics the structure of the building thermal dynamics model and leverages penalty methods to model inequality constraints. Additionally, we use constrained matrix parameterization based on the Perron-Frobenius theorem to bound the eigenvalues of the learned network weights. We interpret the stable eigenvalues as dissipativeness of the learned building thermal model. We demonstrate the effectiveness of the proposed approach on a dataset obtained from an office building with 20 thermal zones.

1 Introduction

Energy-efficient buildings are one of the top priorities to sustainably address the global energy demands and reduction of the CO₂ emissions [1, 2]. Advanced control strategies for buildings have been identified as a potential solution with projected energy saving potential up to 28% [3, 4, 5]. The current state of the art approaches in the domain are based on constrained optimal control methods and heavily depend on the mathematical models of the building dynamics [6, 7, 8, 9, 10].

The building thermal behavior is characterized by high-dimensional, nonlinear, and often discontinuous dynamics, for which modeling requires expertise and development time [11, 12, 13]. Moreover, high computational demands and non-differentiability can easily cast the physics-based model as not suitable for efficient gradient-based optimization that is typically used in various applications. Data-driven system identification typically represents a more cost-efficient alternative [14, 15, 16, 17]. However, purely-black box models require a large amount of data and may not generalize well outside the training distribution [18, 19]. On the other hand, identifying accurate and reliable physics-based models with constrained from data remains a challenging task and involves solving difficult non-convex optimization problems [20, 21, 22]. As a consequence, many of the current control-oriented modeling approaches for buildings still rely on crude approximations assuming low-order linear dynamics [23, 24, 25], which may hamper the overall potential control performance [26].

In this paper we show how to train physics-constrained recurrent neural dynamics models tailored to efficiently learn the building thermal dynamics in an end-to-end fashion, with physically coherent generalization, from small datasets. From a deep learning perspective, the presented model is inspired by a family of neural state-space models (SSM) [27, 28, 29, 30, 31], representing structurally modified vanilla RNNs tailored for the modeling of dynamical systems for control. To the author's best knowledge, this is the first combined use of structured recurrent neural architectures with physics-inspired constraints applied to a real-world building thermal dynamics modeling problem.

2 Methods

Building Thermal Dynamics: When developing predictive models for control purposes, one has to balance model complexity, robustness, and accuracy. The typical building envelope dynamics is represented by a model with a graph structure shown in Fig. 1a. Mathematically, the thermal building model is given as the following differential equation with nonlinear input and disturbance dynamics:

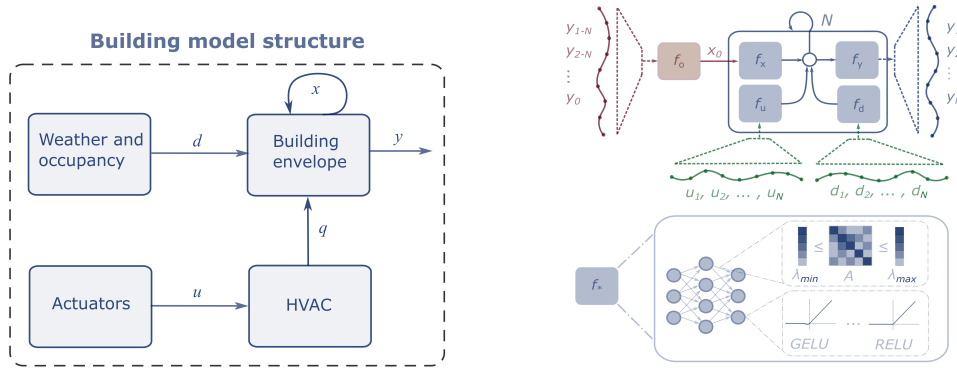
$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + f_d(\mathbf{d}_t), \quad (1a)$$

$$\mathbf{y}_t = C\mathbf{x}_t, \quad (1b)$$

$$\mathbf{u}_t = \dot{\mathbf{m}}_t cp \Delta \mathbf{T}_t, \quad (1c)$$

where \mathbf{x}_t and \mathbf{y}_t represent the values of the states (envelope temperatures), and measurements (zone temperatures) at time t , respectively. Disturbances \mathbf{d}_t represent the influence of weather and occupancy behavior. Heat flows delivered to the building u_k are the product of mass flows $\dot{\mathbf{m}}_t$, difference of the supply and return temperatures $\Delta \mathbf{T}_t$, and the specific heat capacity constant cp .

When the model is built with perfect knowledge from first principles, it is physically interpretable. For instance, the A matrix represents 1-D heat transfer between the spatially discretized system states. B matrix defines the temperature increments caused by the convective heat flow (1c) generated by the HVAC system, while f_d captures highly nonlinear thermal dynamics caused by the weather conditions or internal heat gains generated by the occupancy. However, every building represents a unique system with different operational conditions. Therefore, obtaining the parameters of the differential equations (1) from first principles is a time-consuming, impractical task.



(a) Structure of physics-based building thermal model. (b) Structured recurrent neural dynamics model.

Figure 1: Generic structure of physics-inspired recurrent neural dynamics model architecture. Weights of individual neural blocks f_* are parametrized by linear maps with constrained eigenvalues, while component outputs are subject to penalty constraints parametrized by common activation functions.

Structured Recurrent Neural Dynamics Model: Fig. 1b shows the overall architecture of the physics-inspired neural dynamics model for partially observable systems. To further promote physically coherent behavior, the neural component blocks f_* are parametrized by linear maps with constrained eigenvalues and regularized with penalty functions, as explained in the following sections. The block-structured recurrent neural dynamics model is defined as:

$$\mathbf{x}_{t+1} = f_x(\mathbf{x}_t) + f_u(\mathbf{u}_t) + f_d(\mathbf{d}_t) \quad (2a)$$

$$\mathbf{y}_t = f_y(\mathbf{x}_t) \quad (2b)$$

$$\mathbf{x}_0 = f_o([y_{1-N}; \dots; y_0]) \quad (2c)$$

Here f_x , f_u , and f_d represent decoupled neural components of the overall system model, corresponding to state, input, and disturbance dynamics, respectively. We assume only partially observable systems where states \mathbf{x} represent latent dynamics. As a consequence, we need to use state observer given as additional neural component, f_o , encoding a past N -step window of observations \mathbf{y} onto initial state conditions \mathbf{x}_0 . During training, the model is unrolled and trained on an N -step ahead prediction window. The main advantage of the block nonlinear over unstructured black-box state-space model lies in its structure. The decoupling allows us to leverage prior knowledge for imposing structural assumptions and constraints onto individual blocks of the model.

Eigenvalue Constraints: One key physics insight is that building thermal dynamics represents a dissipative system with stable eigenvalues. This inspired us to enforce physically reasonable constraints on the eigenvalues of a model’s weight matrices. We leverage the method based on the Perron-Frobenius theorem, which states that the row-wise minimum and maximum of any positive square matrix defines its dominant eigenvalue’s lower and upper bound, respectively. Guided by this theorem, we can construct a state transition matrix $\tilde{\mathbf{A}}$ with bounded eigenvalues:

$$\mathbf{M} = \lambda_{\max} - (\lambda_{\max} - \lambda_{\min})\sigma(\mathbf{M}') \quad (3a)$$

$$\tilde{\mathbf{A}}_{i,j} = \frac{\exp(\mathbf{A}'_{ij})}{\sum_{k=1}^{n_x} \exp(\mathbf{A}'_{ik})} \mathbf{M}_{i,j} \quad (3b)$$

We introduce a matrix \mathbf{M} which models damping parameterized by the matrix $\mathbf{M}' \in \mathbb{R}^{n_x \times n_x}$. We apply a row-wise softmax to another parameter matrix $\mathbf{A}' \in \mathbb{R}^{n_x \times n_x}$, then elementwise multiply by \mathbf{M} to obtain our state transition matrix $\tilde{\mathbf{A}}$ with eigenvalues lower and upper bounds λ_{\min} and λ_{\max} .

Inequality Constraints via Penalty Methods: Using an optimization strategy known as the penalty method, we can add further constraints to our model such that its variables remain within physically realistic bounds. We enforce this property by applying inequality constraints via penalty functions $p(\mathbf{y})$ for each time step t :

$$p(\mathbf{y}_t, \underline{\mathbf{y}}_t) : \underline{\mathbf{y}}_t \leq \mathbf{y}_t + \mathbf{s}_t^y \cong \mathbf{s}_t^y = \max(0, -\mathbf{y}_t + \underline{\mathbf{y}}_t) \quad (4a)$$

$$p(\mathbf{y}_t, \bar{\mathbf{y}}_t) : \mathbf{y}_t - \mathbf{s}_t^{\bar{y}} \leq \bar{\mathbf{y}}_t \cong \mathbf{s}_t^{\bar{y}} = \max(0, \mathbf{y}_t - \bar{\mathbf{y}}_t) \quad (4b)$$

The constraints lower and upper bounds are given as $\underline{\mathbf{y}}_k$ and $\bar{\mathbf{y}}_k$, respectively. The slack variables \mathbf{s}_k^y and $\mathbf{s}_k^{\bar{y}}$ indicate the magnitude to which each constraint is violated, and we penalize them heavily in the optimization objective by a large weight on these additional terms in the loss function.

Multi-term Loss Function: We optimize the following loss function augmented with regularization and penalty terms to train the recurrent neural model (2) unrolled over N steps:

$$\begin{aligned} \mathcal{L}_{\text{MSE}}(\mathcal{Y}^{\text{ref}}, \mathcal{Y}|\Theta) = & \frac{1}{N} \sum_{t=1}^N \|\mathbf{y}_t^{\text{ref}} - \mathbf{y}_t\|_2^2 + Q_{\text{dx}} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 + \\ & Q_{\text{ineq}}^y \|\mathbf{s}_t^y\|_2^2 + Q_{\text{ineq}}^u \|\mathbf{s}_t^{f_u}\|_2^2 + Q_{\text{ineq}}^d \|\mathbf{s}_t^{f_d}\|_2^2 \end{aligned} \quad (5)$$

The first term of the loss function computes the mean squared error between predicted \mathbf{y} and observed outputs \mathbf{y}^{ref} over N time steps and represents our primary objective. The term $\mathbf{x}_t - \mathbf{x}_{t-1}$ represents state difference penalty promoting learning of smoother and physically more plausible state trajectories. The violations of the inequality constraints defining the boundary conditions of outputs \mathbf{y} , are penalized by incorporating weighted slack variables \mathbf{s}^y . Thanks to the block-structured dynamics, we can constrain the dynamical contribution of inputs f_u and disturbances f_d towards the overall dynamics via two additional terms in the loss function. This allows us to limit the effect of the external factors to be bounded within physically plausible ranges.

3 Experimental Case Study

Dataset and Experimental Setup The objective is to develop a control-oriented thermal dynamics model of a commercial office building, given a limited amount of measurement data. The time series dataset D is given in the form of tuples with input, disturbance, and output variables, respectively.

$$D = \{(\mathbf{u}_t^{(i)}, \mathbf{d}_t^{(i)}, \mathbf{y}_t^{(i)}), (\mathbf{u}_{t+\Delta}^{(i)}, \mathbf{d}_{t+\Delta}^{(i)}, \mathbf{y}_{t+\Delta}^{(i)}), \dots, (\mathbf{u}_{t+N\Delta}^{(i)}, \mathbf{d}_{t+N\Delta}^{(i)}, \mathbf{y}_{t+N\Delta}^{(i)})\}, \quad (6)$$

where $i = \mathbb{N}_1^n$ represents index of n batches of time series trajectories with N -step horizon. The data is sampled with sampling time $\Delta = 15$ min. We have in total $n_y = 20$ output variables corresponding to zone temperatures, $n_u = 40$ input variables representing HVAC temperatures and mass flows, and $n_d = 1$ disturbance variable for ambient temperature forecast. We use min-max normalization to scale all variables between $[0, 1]$. The dataset consists of 30 days, which corresponds to only 2880 datapoints. We group the dataset into evenly split training, development, and test sets, 960 data points

each. We implement the models using Pytorch [32], and train with randomly initialized weights using the Adam optimizer [33] with a learning rate of 0.003, and 5,000 gradient descent updates. We select the best performing model on the development set, and report results on the test set. The state estimator f_o is a fully connected neural network, while neural blocks f_* are represented by recurrent neural networks with 2 layers and 80 nodes. We range the prediction horizon as powers of two 2^n with $n = 3, \dots, 6$, which corresponds to 2 up to 16 hour window. The relative weights of the multi-term loss function are $Q_{dx} = 0.2$, $Q_{ineq}^y = 1.0$, $Q_{ineq}^u = 0.2$, and $Q_{ineq}^d = 0.2$. We set $\lambda_{min} = 0.8$ and $\lambda_{max} = 1.0$ for stability and dissipativity of learned dynamics.

Results: Fig. 5 assess the simulated open-loop and N -step MSE performance of the recurrent model with and without physics-constraints and structure. The denormalized performance of best-performing models is compared in Tab.2. We observe that imposing physics-inspired structure and constraints not only yields 15% reduction of error but allows us to train models with a larger prediction horizon N . The open-loop MSE of the best-performing constrained and structured model corresponds to 0.488K. In comparison, the state of the art gray-box system identification methods trained on a similar amount of data reports open-loop MSE roughly equal to 1.0K [20]. Hence our preliminary results show more than 100% improvement against state of the art in literature. We also demonstrate the capability to generalize complex dynamics over 30-days from a small dataset of 10-days. For further results analysis, see the appendix with the display of open-loop trajectories and the effect of the eigenvalue constraints.

Table 1: Test set MSE of best constrained structured, and unconstrained unstructured model.

Structure	Constrained	N	N -step [K]	Open-loop [K]
Structured	Y	64	0.4811	0.4884
Unstructured	N	16	0.5266	0.5596

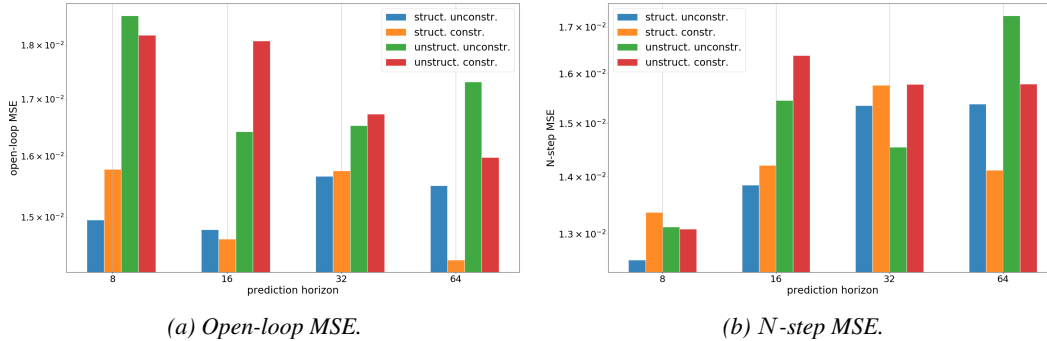


Figure 2: Comparison of open-loop and N -step ahead MSE evaluated on a test set using structured and unstructured models with and without constraints.

4 Conclusions

Reliable data-driven methods which are cost effective in terms of computational demands, data collection, and domain expertise have the potential to revolutionize the field of energy-efficient building operations through wide-scale acquisition of building specific, scalable, and accurate prediction models. We presented a constrained deep learning method for sample-efficient and physics-consistent data-driven modeling of building thermal dynamics. Our approach does not require the large time investments by domain experts and extensive computational resources demanded by physics-based emulator models. Based on only 10 days' measurements, we greatly improve on prior state-of-the-art results for a modeling task using a real-world large scale office building dataset. A potential limitation of the presented approach is the restrictiveness of the used constraints, where wrong initial guess of the eigenvalue and penalty constraints bounds may lead to decreased accuracy of the learned model. Future work includes a systematic comparison against physics-based emulator models and other standard data-driven methods. Authors also plan to use the method as part of advanced predictive control strategies for energy-efficient operations in real-world buildings.

ACKNOWLEDGEMENT

This work was funded by the Physics Informed Machine Learning (PIML) investment at the Pacific Northwest National Laboratory (PNNL). This work emerged from the IBPSA Project 1, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA).

References

- [1] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling climate change with machine learning. *CoRR*, abs/1906.05433, 2019.
- [2] IEA International Energy Agency and International Partnership for Energy Efficiency Cooperation. Building energy performance metrics - supporting energy efficiency progress in major economies. Technical report, IEA Publications, 2015.
- [3] K. W. Roth, D. Westphalen, J. Dieckmann, S. D. Hamilton, and W. Goetzler. Energy Consumption Characteristics of Commercial Building HVAC Systems - Volume III: Energy Savings Potential. Technical report, 2002.
- [4] D. Gyalistras, M. Gwerder, F. Schildbach, C.N. Jones, M. Morari, B. Lehmann, K. Wirth, and V. Stauch. Analysis of Energy Savings Potentials for Integrated Room Automation. In *Clima - RHEVA World Congress*, Antalya, Turkey, May 2010.
- [5] Ján Drgoňa, Javier Arroyo, Iago Cupeiro Figueroa, David Blum, Krzysztof Arendt, Donghun Kim, Enric Perarnau Ollé, Juraj Oravec, Michael Wetter, Draguna L. Vrabie, and Lieve Helsen. All you need to know about model predictive control for buildings. *Annual Reviews in Control*, 2020.
- [6] Jan Široký, Frauke Oldewurtel, Jiří Cigler, and Samuel Prívara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88(9):3079–3087, 2011.
- [7] Y. Ma, F. Borrelli, B. Hancey, B. Coffey, S. Bengea, and P. Haves. Model predictive control for the operation of building cooling systems. *IEEE Transactions on Control Systems Technology*, 20(3):796–803, 2012.
- [8] Mehdi Maasoumy, Barzin Moridian, Meysam Razmara, Mahdi Shahbakhti, and Alberto Sangiovanni-Vincentelli. Online simultaneous state estimation and parameter adaptation for building predictive control. In *ASME 2013 Dynamic Systems and Control Conference*, pages V002T23A006–V002T23A006. American Society of Mechanical Engineers, 2013.
- [9] Vikas Chandan and Andrew G Alleyne. Decentralized predictive thermal control for buildings. *Journal of Process Control*, 24(6):820–835, 2014.
- [10] Simone Baldi, Shuai Yuan, Petr Endel, and Ondrej Holub. Dual estimation: Constructing building energy models from data sampled at low rate. *Applied Energy*, 169:81–92, 2016.
- [11] F. Jorissen, G. Reynders, R. Baetens, D. Picard, D. Saelens, and L. Helsen. Implementation and verification of the IDEAS building energy simulation library. *Journal of Building Performance Simulation*, 11(6):669–688, 2018.
- [12] M. Wetter, W. Zuo, T. Nouidui, and X. Pang. Modelica buildings library. *Journal of Building Performance Simulation*, 7(4):253–270, 2014.
- [13] Hao Gao, Christian Koch, and Yupeng Wu. Building information modelling based building energy modelling: A review. *Applied Energy*, 238:320 – 343, 2019.

- [14] Nadia D. Roman, Facundo Bre, Victor D. Fachinotti, and Roberto Lamberts. Application and characterization of metamodels based on artificial neural networks for building performance simulation: A systematic review. *Energy and Buildings*, 217:109972, 2020.
- [15] Yuna Zhang, Zheng O'Neill, Bing Dong, and Godfried Augenbroe. Comparisons of inverse modeling approaches for predicting building energy performance. *Building and Environment*, 86:177 – 190, 2015.
- [16] Francesco Smarra, Achin Jain, Tullio de Rubeis, Dario Ambrosini, Alessandro D'Innocenzo, and Rahul Mangharam. Data-driven model predictive control using random forests for building energy optimization and climate control. *Applied Energy*, 226:1252 – 1272, 2018.
- [17] L. Ferkl and J. Šíroký. Ceiling radiant cooling: Comparison of armax and subspace identification modelling methods. *Building and Environment*, 45(1):205 – 212, 2010.
- [18] Abdul Afram and Farrokh Janabi-Sharifi. Review of modeling methods for HVAC systems. *Applied Thermal Engineering*, 67(1):507 – 519, 2014.
- [19] Zakia Afroz, GM Shafiullah, Tania Urmee, and Gary Higgins. Modeling techniques used in building HVAC control systems: A review. *Renewable and Sustainable Energy Reviews*, 2018.
- [20] Javier Arroyo, Fred Spiessens, and Lieve Helsen. Identification of multi-zone grey-box building models for use in model predictive control. *Journal of Building Performance Simulation*, 13(4):472–486, 2020.
- [21] Krzysztof Arendt, Muhyiddine Jradi, Michael Wetter, and Christian Veje. ModestPy: An Open-Source Python Tool for Parameter Estimation in Functional Mock-up Units. In Michael Tiller, Hubertus Tummescheit, and Luigi Vanfretti, editors, *Proceedings of the 1st American Modelica Conference*, pages 121–130. Modelica Association and Linköping University Electronic Press, 10 2018.
- [22] Hao Huang, Lei Chen, and Eric Hu. Model predictive control for energy-efficient buildings: An airport terminal building study. *IEEE International Conference on Control and Automation, ICCA, Taichung, Taiwan*, pages 1025–1030, 06 2014.
- [23] Samuel Prívara, Jiří Cigler, Zdeněk Váňa, Frauke Oldewurtel, and Eva Žáčková. Use of partial least squares within the control relevant identification for buildings. *Control Engineering Practice*, 21(1):113 – 121, 2013.
- [24] Glenn Reynders, Jan Diriken, and Dirk Saelens. Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals. *Energy and Buildings*, 82:263–274, 2014.
- [25] T. Zakula, P.R. Armstrong, and L. Norford. Modeling environment for model predictive control of buildings. *Energy and Buildings*, 85:549 – 559, 2014.
- [26] D. Picard, J. Drgoňa, M. Kvasnica, and L. Helsen. Impact of the controller model complexity on model predictive control performance for buildings. *Energy and Buildings*, 152:739 – 751, 2017.
- [27] Rahul G. Krishnan, Uri Shalit, and David Sontag. Structured inference networks for nonlinear state space models. *AAAI*, 2016.
- [28] Danijar Hafner, Timothy P. Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. *CoRR*, abs/1811.04551, 2018.
- [29] D. Masti and A. Bemporad. Learning nonlinear state-space models using deep autoencoders. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 3862–3867, 2018.
- [30] Syama S. Rangapuram, Matthias W. Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 7785–7794. Curran Associates, Inc., 2018.

- [31] Olalekan P. Ogunmolu, Xuejun Gu, Steve B. Jiang, and Nicholas R. Gans. Nonlinear systems identification using deep dynamic neural networks. *CoRR*, abs/1610.01439, 2016.
- [32] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8024–8035, 2019.
- [33] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [34] Javier Rubio-Herrero, Vikas Chandan, Charles Siegel, Abhinav Vishnu, and Draguna Vrabie. A learning framework for control-oriented modeling of buildings. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 473–478. IEEE, 2017.
- [35] Dan Hendrycks and Kevin Gimpel. Bridging nonlinearities and stochastic regularizers with gaussian error linear units. *CoRR*, abs/1606.08415, 2016.
- [36] Alice Mugnini, Gianluca Coccia, Fabio Polonara, and Alessia Arteconi. Performance assessment of data-driven and physical-based models to predict building energy demand in model predictive controls. *Energies*, 13(12), 2020.
- [37] Damien Picard, Maarten Sourbron, Filip Jorissen, Jiri Cigler, Lukás Ferkl, and Lieve Helsen. Comparison of model predictive control performance using grey-box and white-box controller models. In *Proceedings of the 4th International High Performance Buildings Conference, West Lafayette, IN, USA*, pages 1–10, West-Lafayette, Indiana, USA, 2016.
- [38] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. Understanding the exploding gradient problem. *CoRR*, abs/1211.5063, 2012.
- [39] J. F. Kolen and S. C. Kremer. *Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies*, pages 237–243. 2001.
- [40] P. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2008.
- [41] Jonathan H. Tu, Clarence W. Rowley, Dirk M. Luchtenburg, Steven L. Brunton, and J. Nathan Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1(2158-2491), 2014.

A Extended Results

The objective is to develop a control-oriented model of the thermal dynamics of a commercial office building, given only a limited amount of time series measurement data.

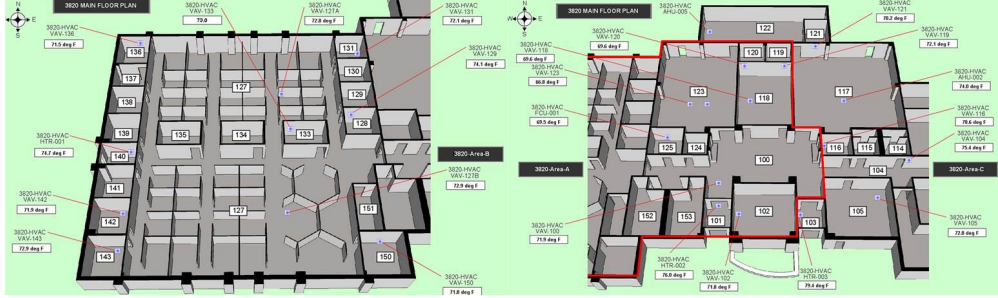
A.1 Real-world Building Dataset and Experimental Setup

Real-world Building Dataset: The building used in this study is a commercial building in Richland, WA described in [34]. Heating and cooling are provided by a variable air volume (VAV) system served by 4 air handling units (AHUs) serving 24 VAV boxes (zones). Each VAV box is equipped with a hot water reheat coil. A boiler, fed by natural gas, supplies hot water to the reheat coils and AHU coils. Chilled water is supplied by a central chiller plant.

Data from specific sensors for the above-mentioned buildings is stored in a database, which communicates with the building management system (BMS) and polls data for these sensors at a time resolution of 1 minute. A total of more than 600 sensors report data corresponding to measurements such as supply and return temperatures of air and water, air, hot water and cold water flow rates, energy and power consumption, set-points for the underlying control systems, occupancy status in zones, and outside air temperature. Data was cleaned and pre-processed according to the methodology described in [34]. The same dataset was used in [34] to model the building’s power consumption and zone temperatures using RNN model with LSTM architecture. The authors have been able to achieve high



(a) Building's facade.



(b) Building's zone layout.

Figure 3: Office building of interest.

prediction accuracy on a single step ahead prediction compared to other standard machine learning models such as linear regression, support vector regression, and random forests. However, due to the purely black-box nature and 1-step ahead loss function, the model in [34] does not explicitly guarantee physical constraints and is not suitable for long-term predictions of the building's thermal behavior. This case study demonstrates improved accuracy, generalization, long-term prediction capabilities, and physically coherent and interpretable dynamic behavior of the learned dynamical model with 20 thermal zones. Hence, considering a model with higher complexity compared to the 2-zone model presented in [34].

Experimental Setup: We implement the presented model architectures using Pytorch [32], and train with randomly initialized weights using the Adam optimizer [33] with a learning rate of 0.003, and 5,000 gradient descent updates. We select the best performing model on the development set from a directed hyperparameter search. All neural network blocks are designed with GELU activation functions [35]. The state estimator is encoded with a fully connected neural network, while individual neural blocks f_* are represented either by standard multilayer perceptron (MLP), recurrent neural network (RNN), or residual neural network (ResNet), respectively, each with 2 layers and 80 nodes. We range the prediction horizon as powers of two 2^n with $n = 3, \dots, 6$, which corresponds to 2 up to 16 hour prediction window. The relative weights of the multi-term loss function for constrained models are $Q_{dx} = 0.2$, $Q_{ineq}^y = 1.0$, $Q_{ineq}^u = 0.2$, and $Q_{ineq}^d = 0.2$. We set $\lambda_{\min} = 0.8$ and $\lambda_{\max} = 1.0$ for stability and low dissipativity of learned dynamics when using eigenvalue constraints.

A.2 Results and Analysis

This section assesses the open-loop and N -step simulation performance of trained recurrent neural dynamics models with and without structure and constraints, respectively. We systematically compare and analyze the added value of the block structure, penalty, and eigenvalue constraints, where Tab. 2 summarizes the best performance of the modeling variants. Moreover, we discuss the interpretability of the proposed data-driven models through the optics of building physics.

Best Performing Model: As reported in Tab. 2, we achieve the best performance with constrained and structured recurrent neural model (2). The best model scores 0.0052, 0.0091, and 0.0143, on

Table 2: Test set MSE of best-performing structured, unstructured, constrained and unconstrained models, respectively.

Structure	Constrained	Weights	N	N -step [K]	Open-loop [K]
Structured	Y	Linear	64	0.4811	0.4884
	N	Perron-Frobenius	16	0.4720	0.5043
Unstructured	Y	Linear	64	0.5380	0.5446
	N	Linear	16	0.5266	0.5596

normalized open-loop MSE evaluated on the test, dev, and train set, respectively. From a physical perspective, the denormalized open-loop MSE corresponds to roughly 0.18K, 0.31K, and 0.49K errors per output, respectively. This demonstrates the ability to generalize the dynamics over the period of 30 days, given only 10 days of training data. In comparison, the state of the art gray-box and black-box system identification methods trained on a similar amount of data reports open-loop MSE greater than 1.0K [20, 36, 37]. Hence our results show more than 100% improvement against state of the art. However, a more rigorous comparison needs to be performed to compare the prediction errors with standard gray-box methods using the same datasets. For visual assessment, Fig. 4 shows normalized open-loop simulation trajectories of best performing structured dynamics model on the train, dev, and test set, represented by gray zones, respectively.

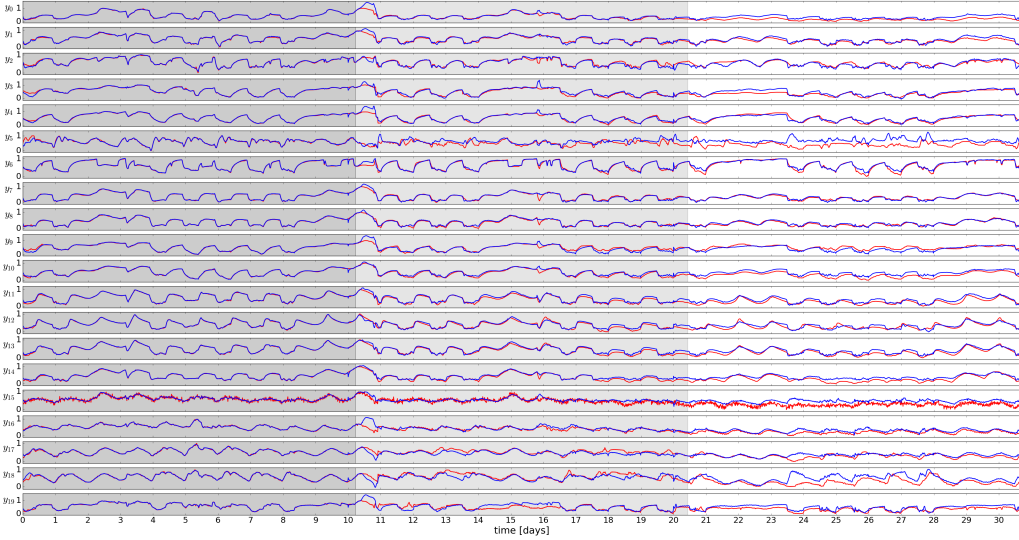


Figure 4: Open-loop trajectories of the learned (blue) and ground truth (red) multi-zone building thermal dynamics.

Effect of Prediction Horizon and Penalty Constraints: Fig. 5 shows test set performance with open-loop MSE and N -step ahead MSE losses for structured and constrained model variants trained with increasing prediction horizon N . As expected, Fig. 5b shows that N -step MSE rises with a longer prediction horizon in the training loss function because learning long-term predictions is generally a more difficult task. Tab. 2 reports larger MSE gaps between N -step and open-loop loss for smaller prediction horizon $N = 16$. On the other hand, larger horizon $N = 64$ minimizes the gap between N -step loss function and open-loop performance, hence providing a more accurate assessment of the desired performance measure. Also, as shown in Fig. 5a, longer prediction horizon tends to improve the overall open-loop simulation performance of all constrained models. The same does not hold for unconstrained models for which the performance starts to deteriorate with a horizon longer than 16. This indicates that including penalty constraints in the training loss function helps to improve the model accuracy over longer prediction horizons. The intuition here is simple; by confining the system outputs into a physically meaningful subspace, the model is less likely to learn diverging long-term trajectories.

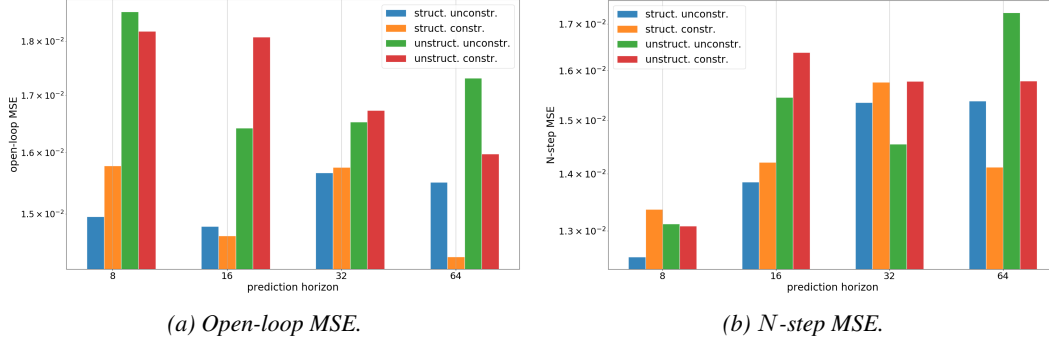


Figure 5: Effect of penalty constraints on open-loop and N -step ahead MSE evaluated on a test set using structured and unstructured models, with increasing training prediction horizon N .

Effect of Physics-inspired Structure: Fig. 5 demonstrates that adding building-physics inspired structure into the neural state-space model undeniably improves both open-loop and N -step MSE. Results in Tab. 2 confirm that both constraints and structure have a positive influence on the open-loop performance of trained models, while structure being a more significant modeling assumption. Applying both structure and constraints yields a 15% reduction in prediction error against unstructured and unconstrained neural state-space model counterparts. By decoupling the state, control action, and disturbance dynamics into separate blocks modeled by neural networks, we prevent the model from learning lumped dynamics behavior. Each block can now learn different nonlinear transformations, which can be independently interpreted as structural heat transfer dynamics for states f_x , HVAC dynamics for inputs f_u , and weather and occupancy thermal dynamics for disturbance signals f_d .

Effect of Neural Blocks Architecture: Fig. 6 shows the effect on open-loop and N -step MSE of using different neural architectures for representing the individual blocks of structured and unstructured neural state-space models, respectively. We focus our analysis on best-case open-loop performance displayed in Fig. 7a. Please note the y-axis is in the logarithmic scale. Surprisingly, models with ResNet architectures are less accurate than best performing RNN or MLP across all prediction horizons and deteriorate fast with increasing prediction horizon. The cause of ResNets' poor performance is hard to estimate at this point, and more in-depth analysis needs to be performed in the future. On the other hand, the performance of models with both RNN and MLP blocks is comparable and scales well also with larger horizons. While models with RNN blocks tend to perform better for shorter horizons, models with MLP architecture score better for the largest time horizon of 64 steps. This might be linked with well known RNN issues, such as vanishing, and exploding gradient problems causing difficulties when learning long-term dependencies [38, 39].

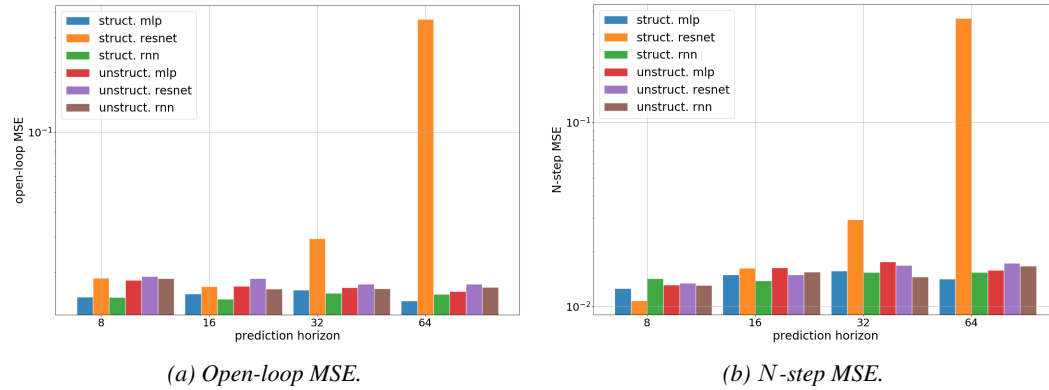


Figure 6: Effect of neural blocks architecture on open-loop and N -step ahead MSE evaluated on a test set using structured and unstructured models, with increasing training prediction horizon N .

Effect of Weight's Eigenvalue Constraints: Fig. 7 shows test set performance with open-loop MSE and N -step ahead MSE losses for structured and unstructured model variants with and without eigenvalue constraints via pf factorization of weights. Due to restrictive nature of the pf factorization, in Fig. 7b we observe larger increase in N -step MSE compared to unconstrained linear weights for most of the cases. However, as shown in in Fig. 7a, the eigenvalue constraints improve the performance of the structured models for shorter prediction horizons, as a consequence of the imposed inductive bias towards learning dissipative heat transfer dynamics. On the other hand, unstructured models do not benefit from using pf factorization at all. The reason is that imposed eigenvalue constraints are inspired by the building envelope dynamics exclusively modeled with f_x map of structured models (2). In contrast, unstructured models learn lumped envelope, HVAC, and disturbance dynamics, hence they fail to benefit from any block-specific priors.

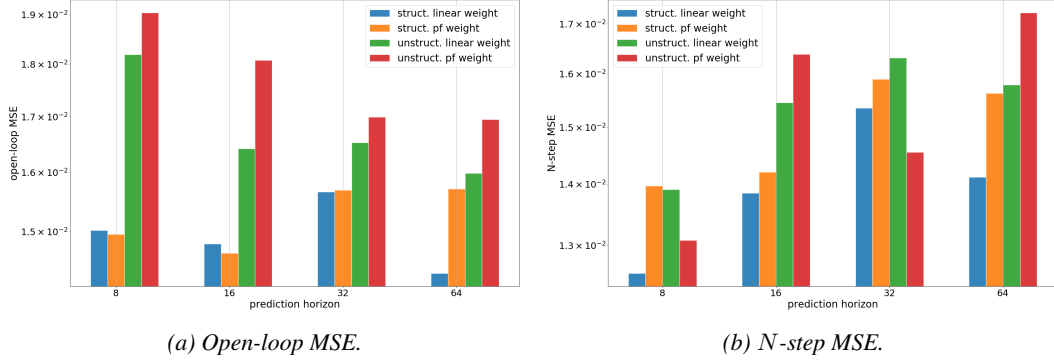


Figure 7: Effect of eigenvalue constraints via pf factorization on open-loop and N -step ahead MSE evaluated on a test set using structured and unstructured models, with increasing training prediction horizon N .

A.3 Eigenvalue Analysis and Physical Interpretability

Fig. 8 shows concatenated eigenvalues in the complex plane for weights of the state transition maps f_x and f of learned structured (2) and unstructured recurrent neural dynamics models, respectively. Besides structure Fig. 8 compares the effect of eigenvalue constraints using Perron-Frobenius (pf) factorization of the system dynamics weights. Please note that we plot only eigenvalues of the neural network's weights. Hence the dynamic effects of the activation functions are omitted in this analysis. However, all our neural network blocks are designed with GELU activation functions, which represent contractive maps with strictly stable eigenvalues. Therefore, based on the argument of the composition of stable functions, the global stability of the learned dynamics is not compromised.

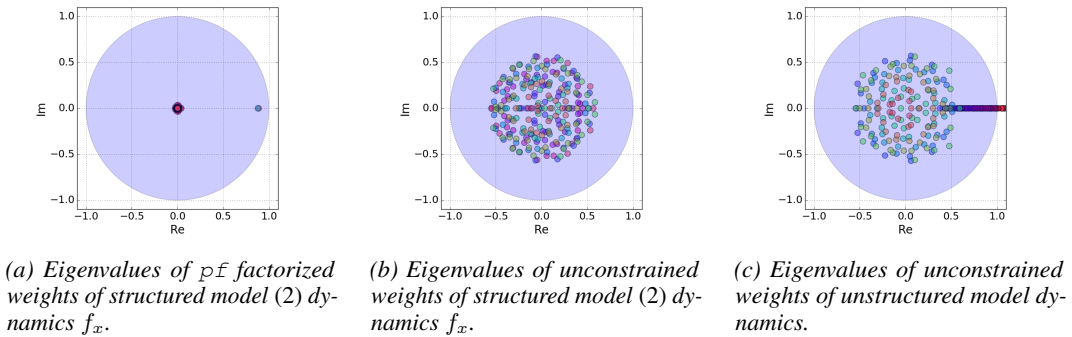


Figure 8: Eigenvalue plots of the weights of system dynamics maps f_x and f of learned structured and unstructured dynamical models, respectively. Blue circles represent stable regions.

Fig. 8a shows the effect of proposed eigenvalue constraints pf factorization, and verifies that the dominant eigenvalue remains within prescribed bounds $\lambda_{\min} = 0.8$ and $\lambda_{\max} = 1.0$. Hence the

dissipativeness of the learned dynamics is hard constrained within physically realistic values when using pf factorization. Another interesting observation is that there are only two dominant dynamical modes with eigenvalues larger than 0.8, one per each layer of f_x . While the rest of the eigenvalues fall within 0.05 radius, hence representing less significant dynamic modes [40, 41]. This indicates a possibility to obtain lower-order representations of the underlying higher-order nonlinear system, a property useful for real-time optimal control applications.

In contrast, as displayed in Fig. 8a and Fig. 8c, the eigenvalues of standard unconstrained weights for both structured and unstructured models are more dispersed with larger imaginary parts. The imaginary parts indicate oscillatory modes of the autonomous state dynamics f_x and f , respectively. However, in the case of building thermal dynamics, the periodicity of the dynamics is caused by external factors such as weather and occupancy schedules. From this perspective, the structured models using pf factorization of the weights, are closer to the physically realistic parameterization of the system dynamics. Additionally, not using eigenvalue constraints may result in learning unstable weights. Fig. 8c displays an example where the unstructured learned model does not guarantee the satisfaction of physically realistic dissipativeness property.