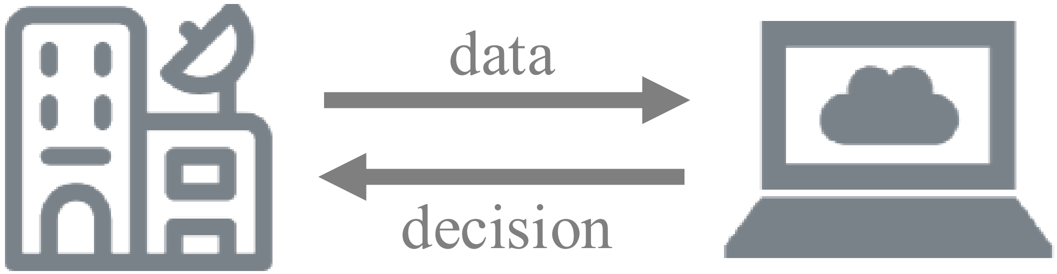
Massachusetts Institute of Technology

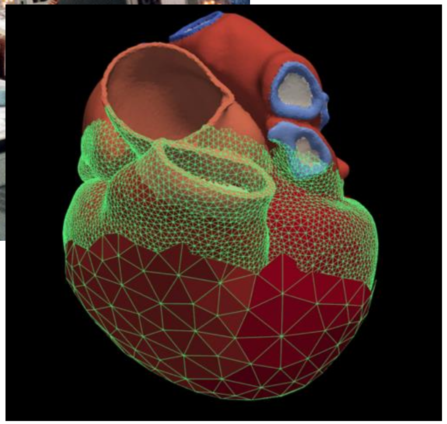
Data-efficient scientific machine learning for urban intelligence and sustainability

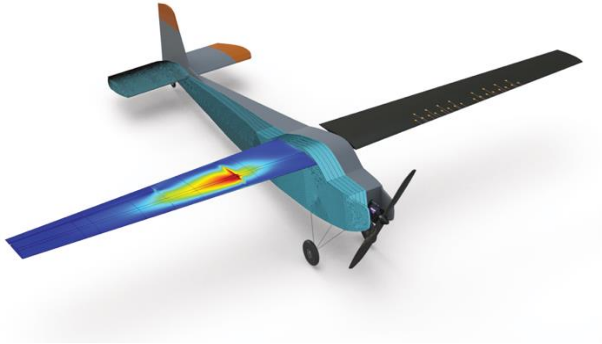
Sicheng Zhan
Postdoctoral associate, MIT Building Technology
25/11/18

Data-centric scientific machine learning for urban intelligence and sustainability

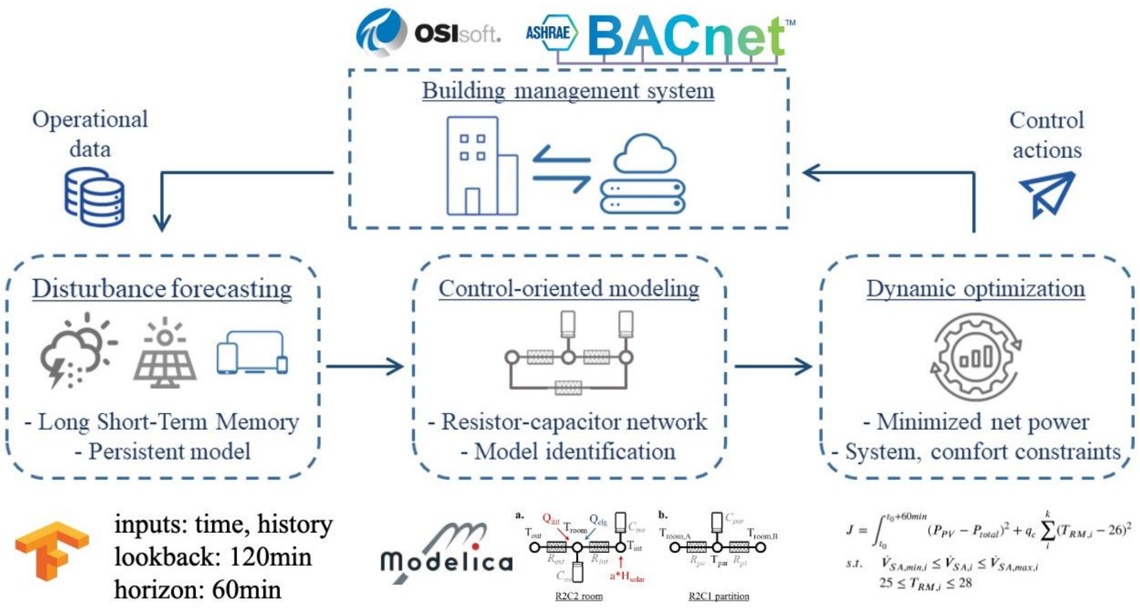
Digital twins - Computational models that replicate the behaviour of real-world systems, conducting virtual experiments in unseen scenarios and supporting decision-making





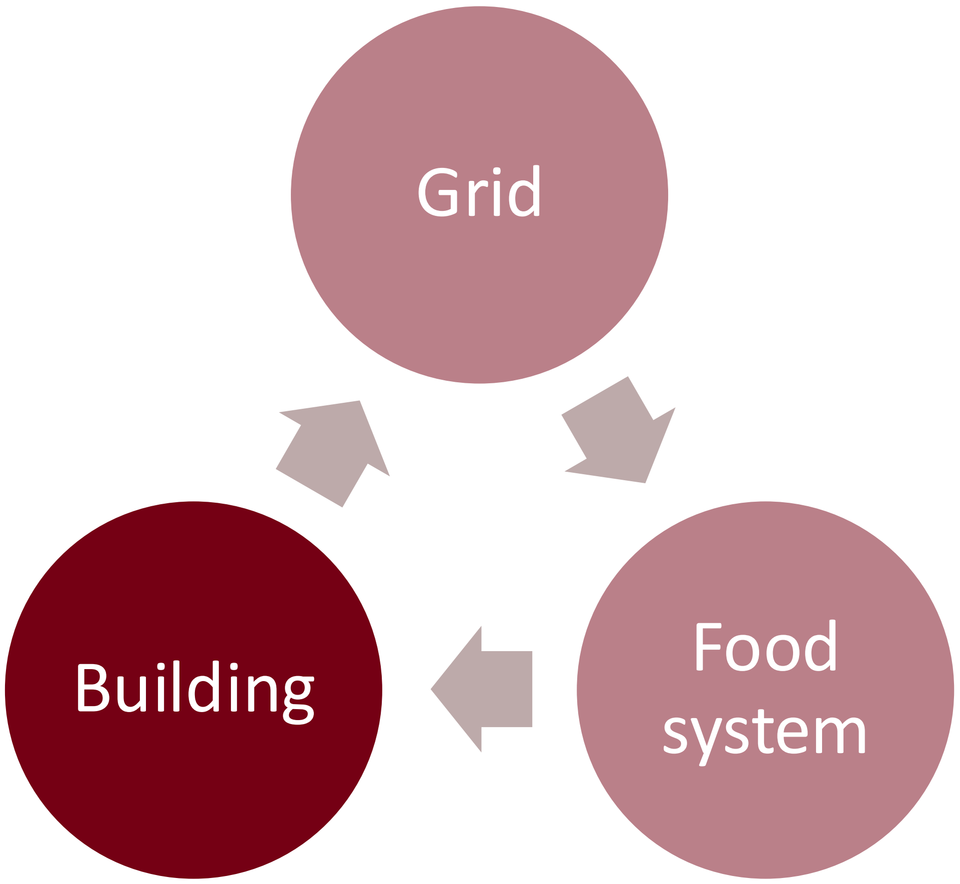


Data-centric scientific machine learning for urban intelligence and sustainability

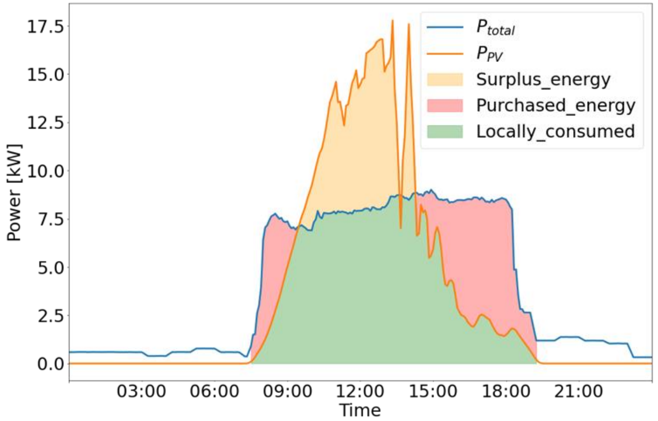


Optimal control implemented at multiple real-world buildings

Thermal comfort, energy conservation, self-sufficiency, etc.



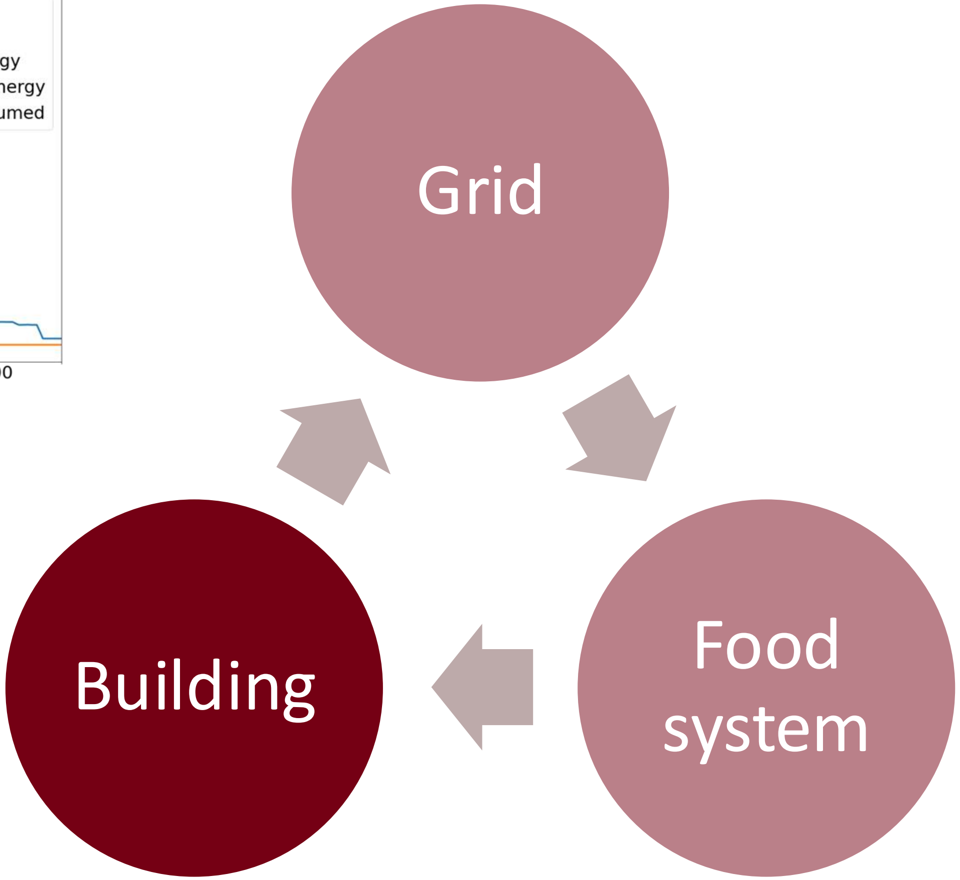
Data-centric scientific machine learning for **urban intelligence and sustainability**



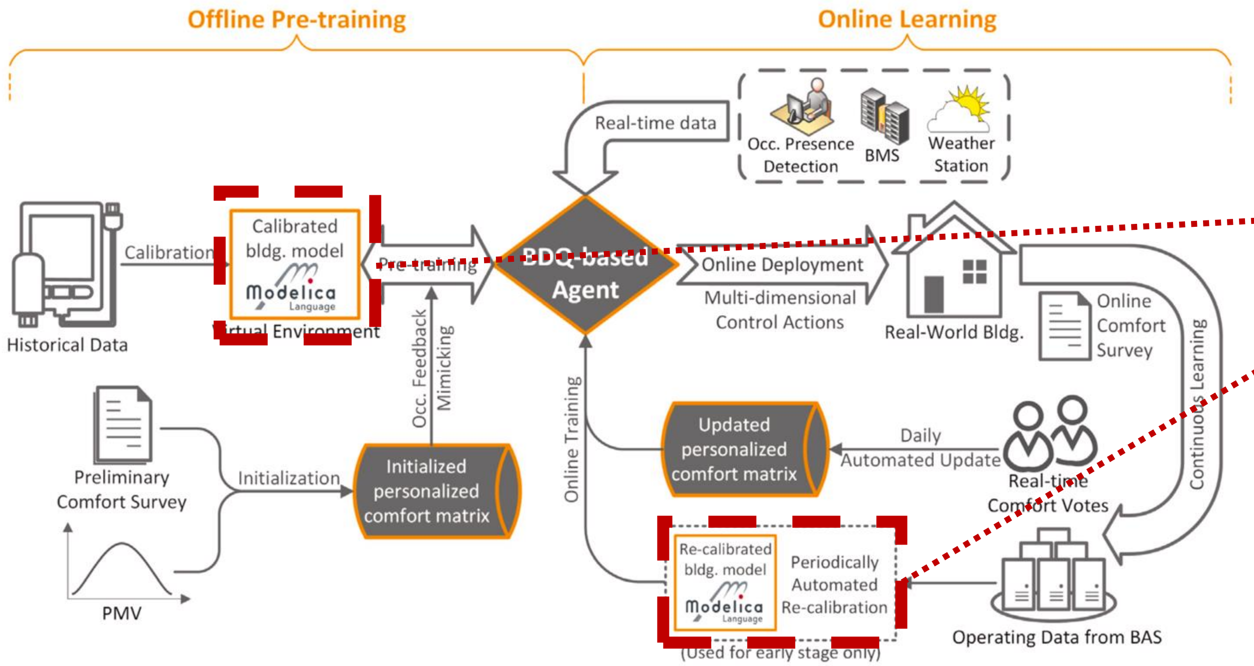
Self-consumption: $$SC = \frac{E_{locally\_consumed}}{E_{PV}}$$

Self-sufficiency: $$SS = \frac{E_{locally\_consumed}}{E_{total}}$$

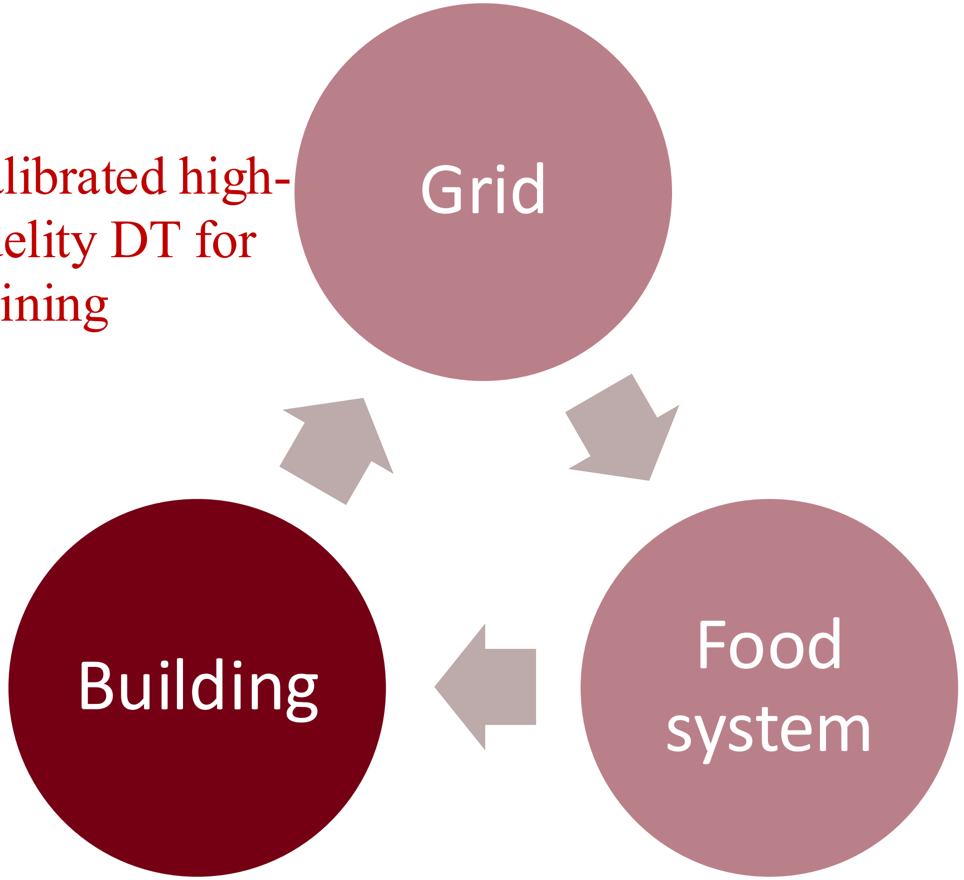
SC improved by 19.5%, SS improved by 10.6%



Data-centric scientific machine learning for urban intelligence and sustainability

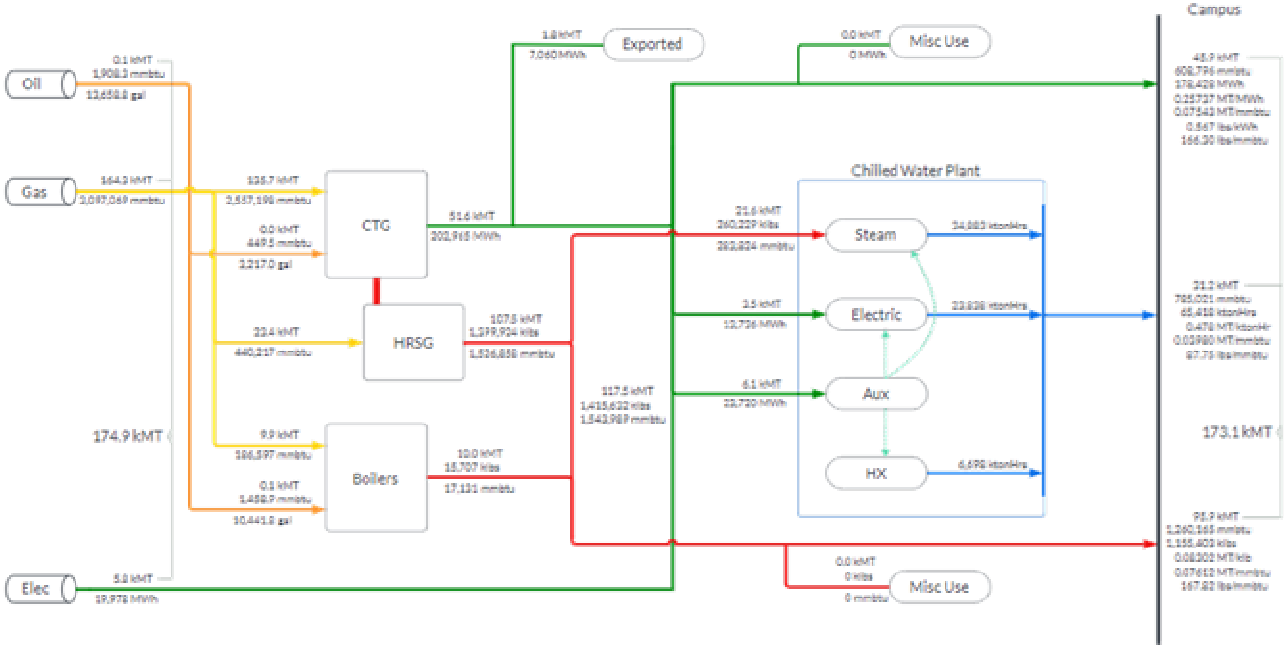


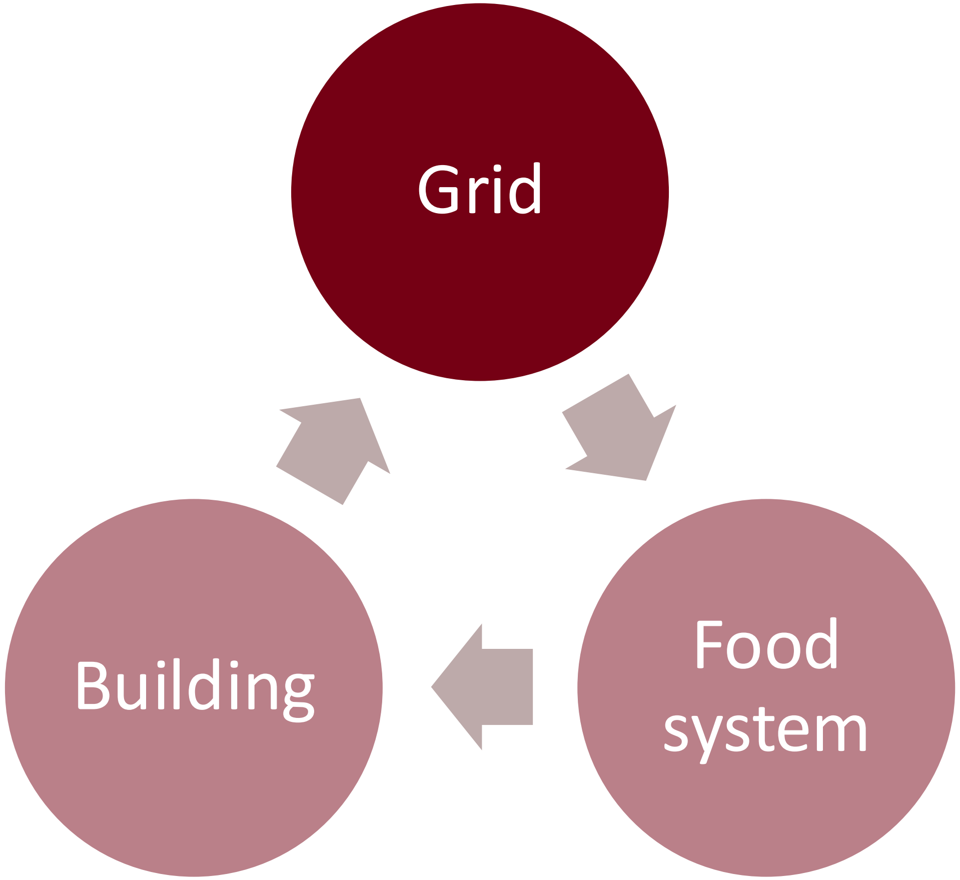
Calibrated high-fidelity DT for training



- Occupant in the loop optimal control of temperature setpoints and fan speeds
- 14% reduction in cooling energy and 11% improvement in total thermal acceptability

Data-centric scientific machine learning for urban intelligence and sustainability

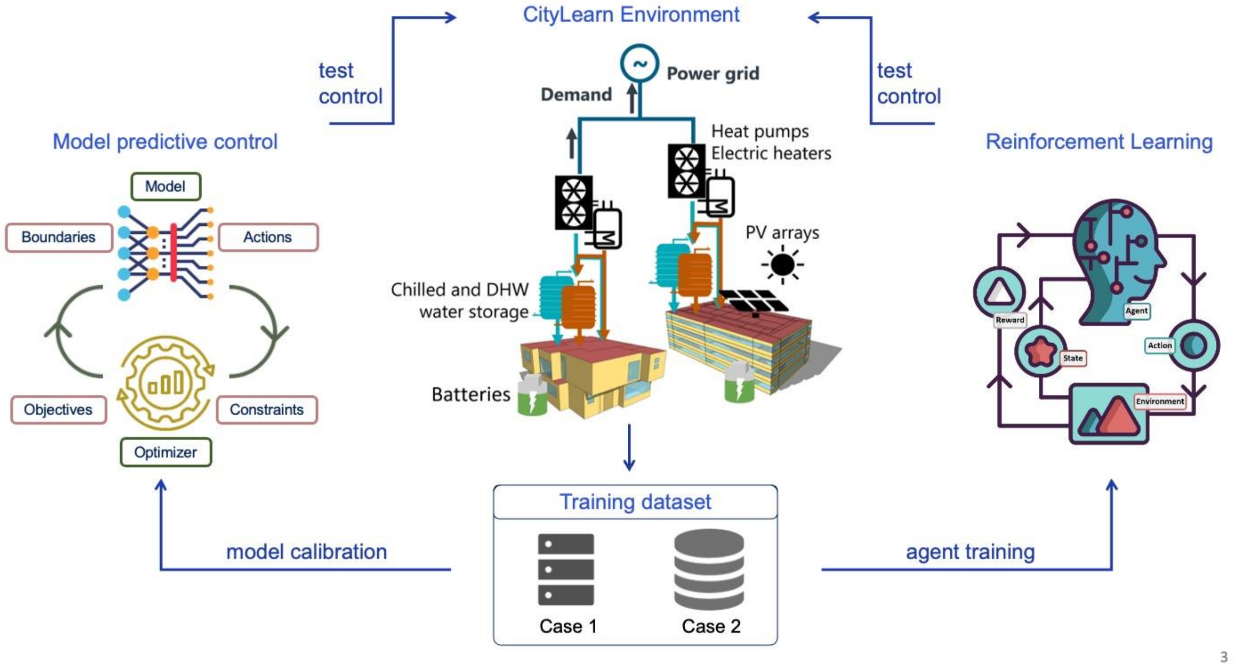


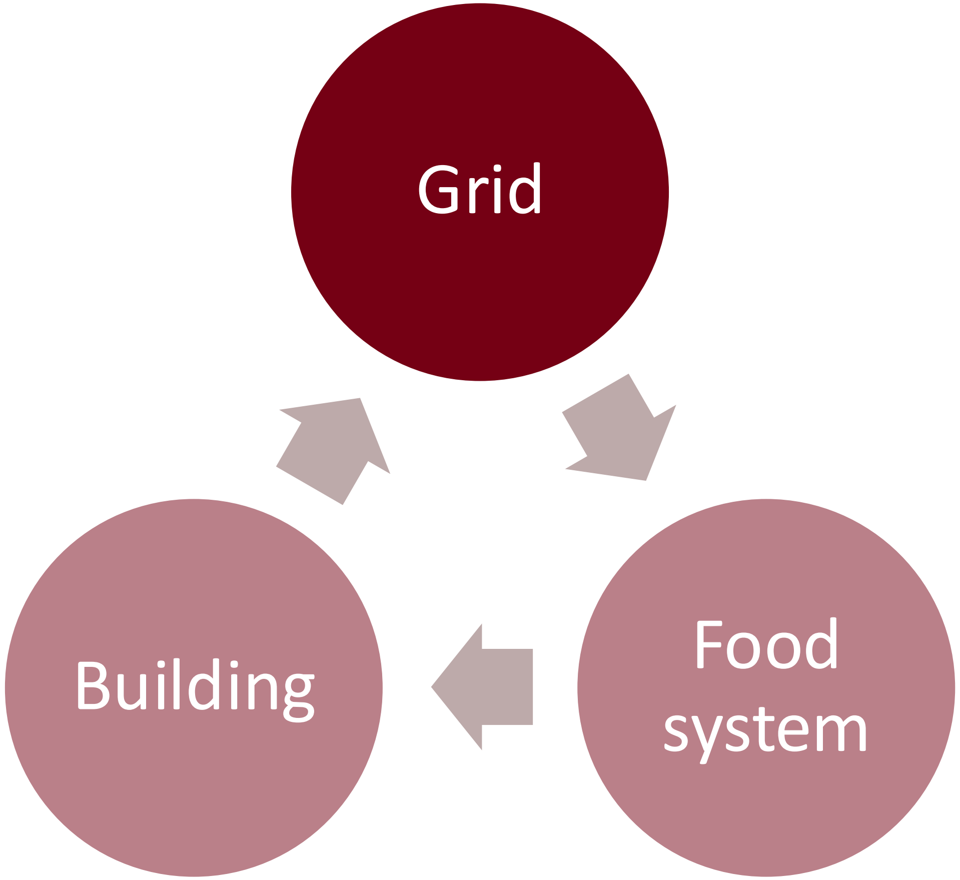


Minimize electricity purchase (total and peak load) from the main grid

Reduce carbon intensity by coordinating electricity, heating, and cooling loads

Data-centric scientific machine learning for urban intelligence and sustainability

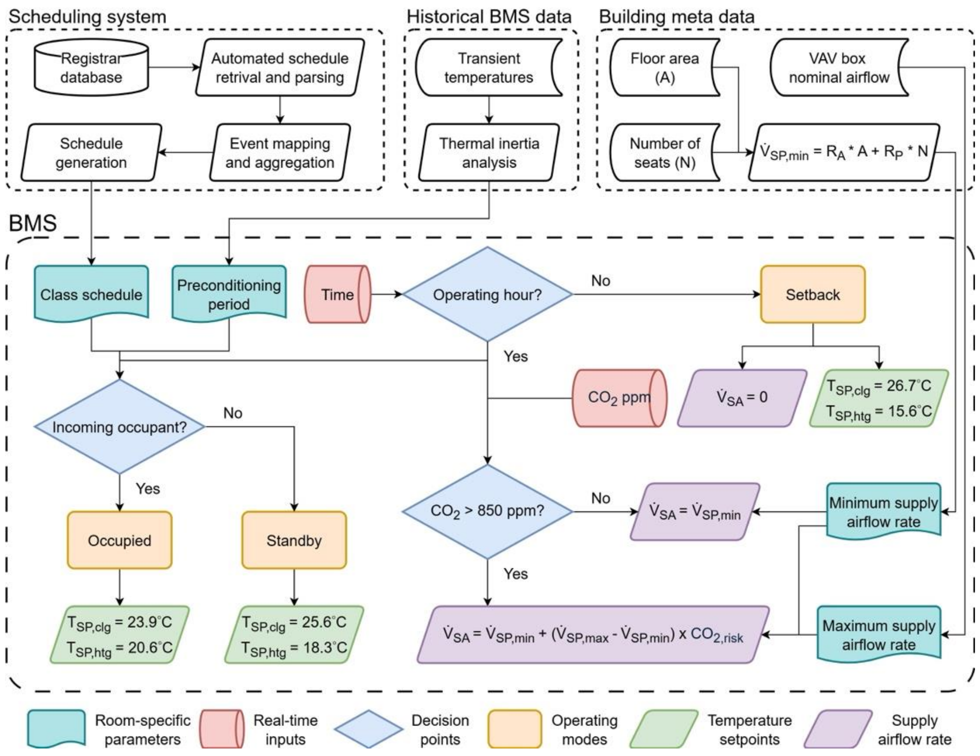




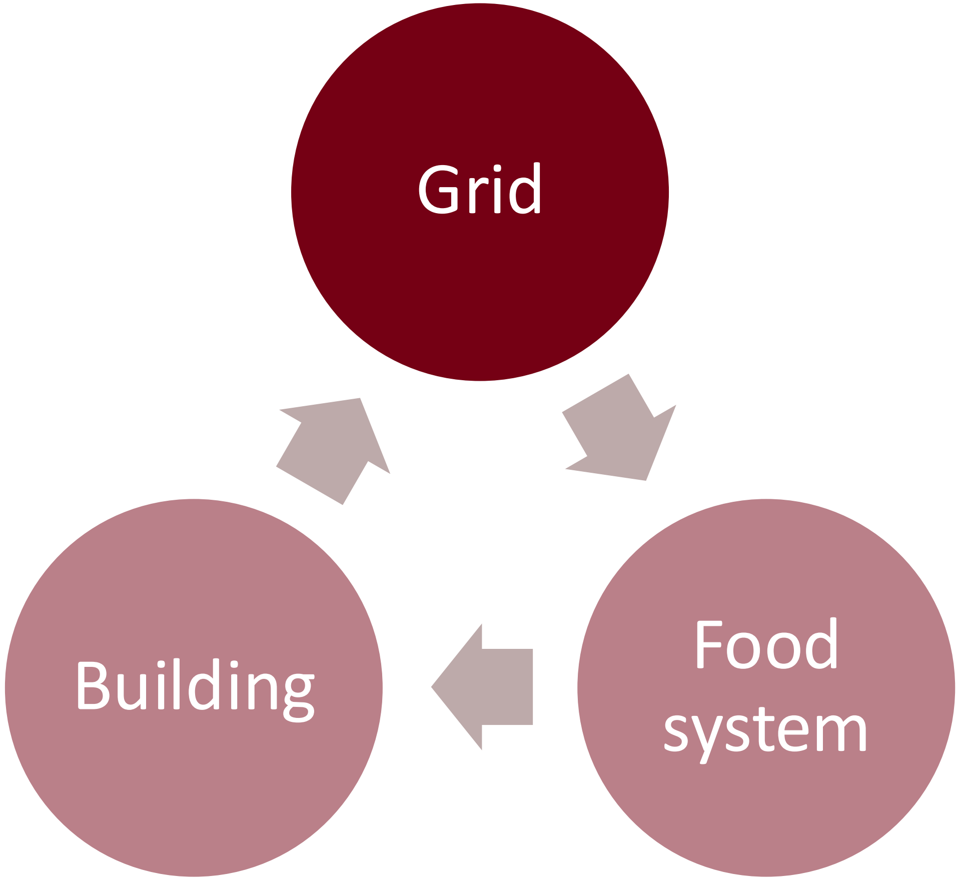
Building-grid interaction, distributed renewable energy, district energy system decarbonization

More scalable solutions required for complex systems and tasks

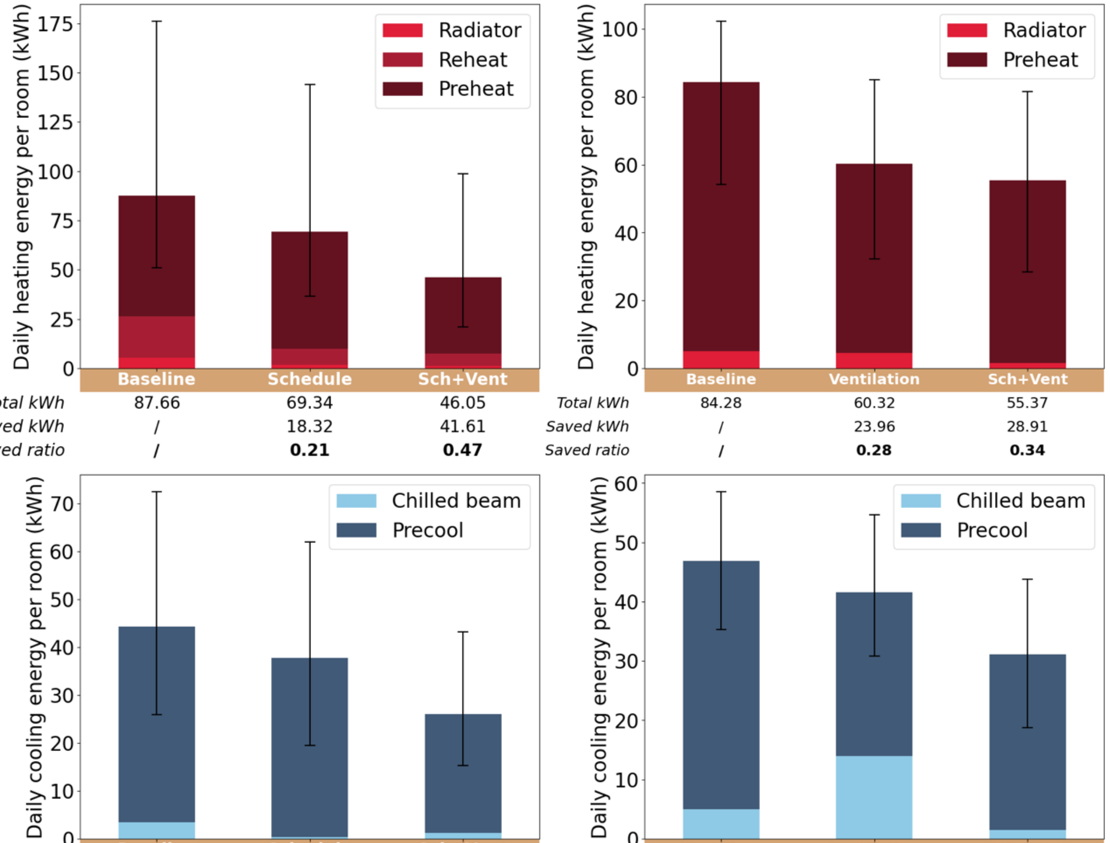
Data-centric scientific machine learning for **urban intelligence and sustainability**



- Rule-based control to obtain near-optimal control actions
- DT for monitoring and measurement & verification
- Over 40% energy saving and ~30% carbon reduction



Data-centric scientific machine learning for urban intelligence and sustainability

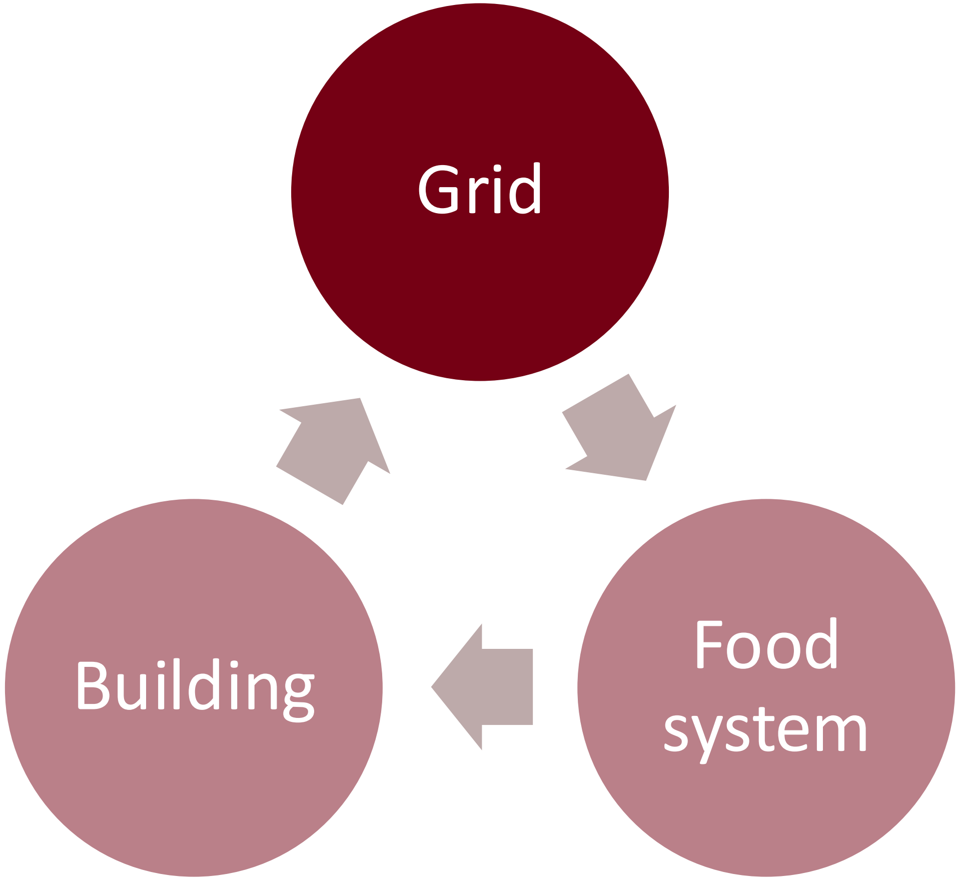


	Baseline	Schedule	Sch+Vent
Total kWh	87.66	69.34	46.05
Saved kWh	/	18.32	41.61
Saved ratio	/	**0.21**	**0.47**

	Baseline	Ventilation	Sch+Vent
Total kWh	84.28	60.32	55.37
Saved kWh	/	23.96	28.91
Saved ratio	/	**0.28**	**0.34**

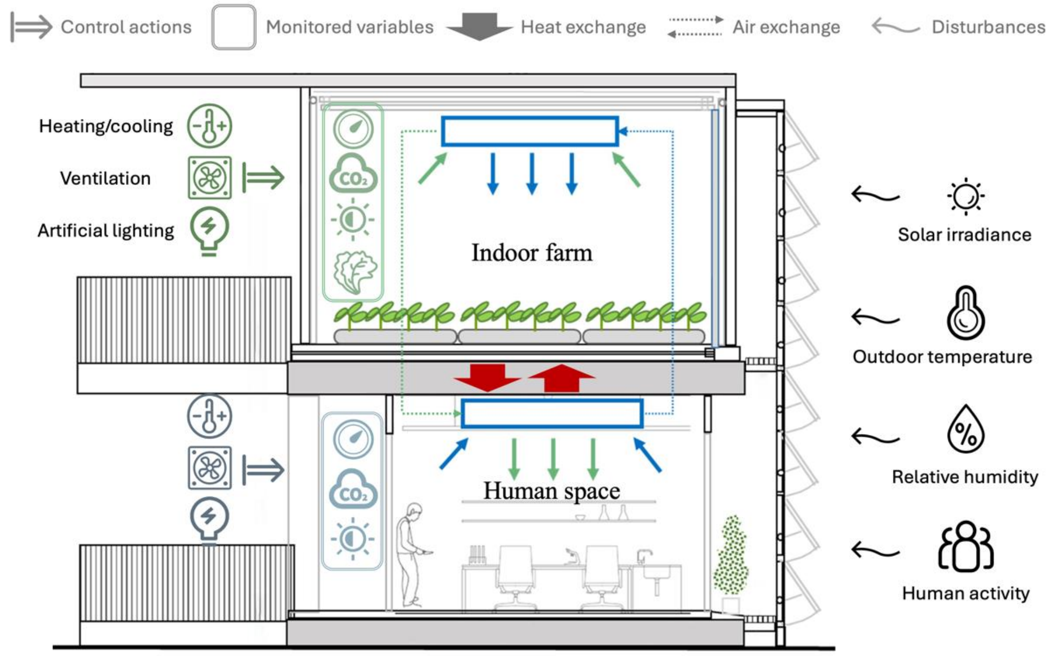
	Baseline	Schedule	Sch+Vent
Total kWh	44.27	37.75	25.98
Saved kWh	/	6.52	18.29
Saved ratio	/	**0.15**	**0.41**

	Baseline	Ventilation	Sch+Vent
Total kWh	46.87	41.6	31.06
Saved kWh	/	5.27	15.81
Saved ratio	/	**0.11**	**0.34**

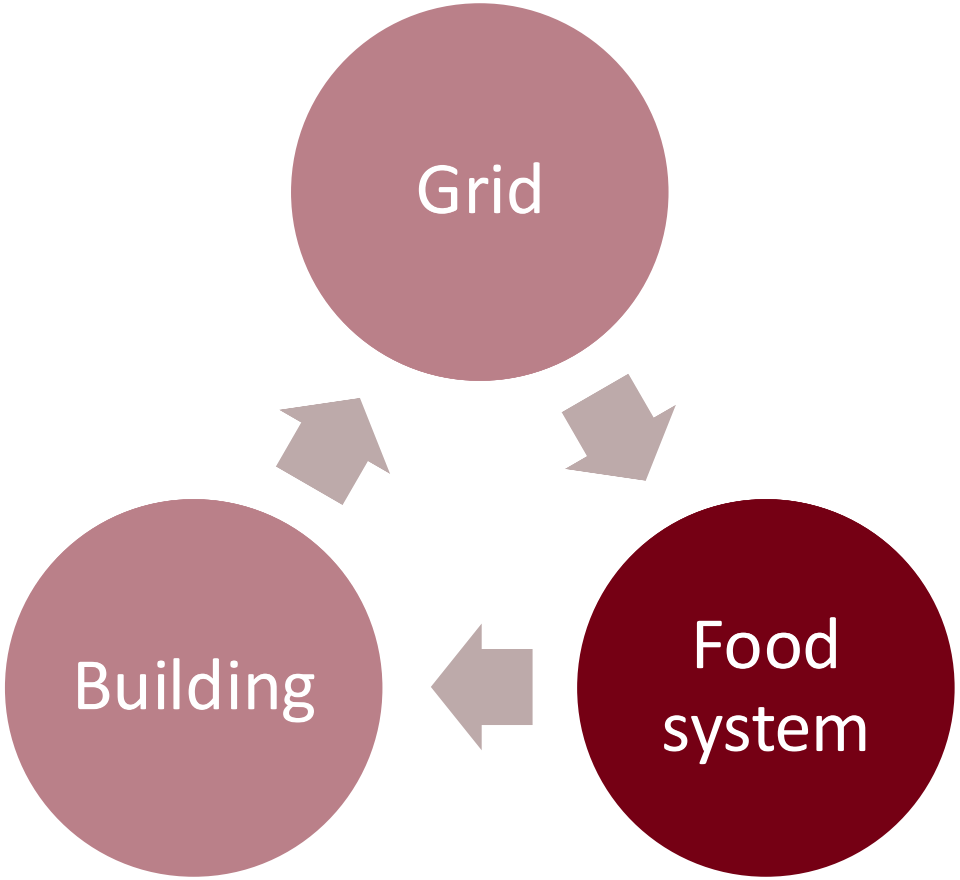


- Over 40% energy saving and ~30% carbon reduction

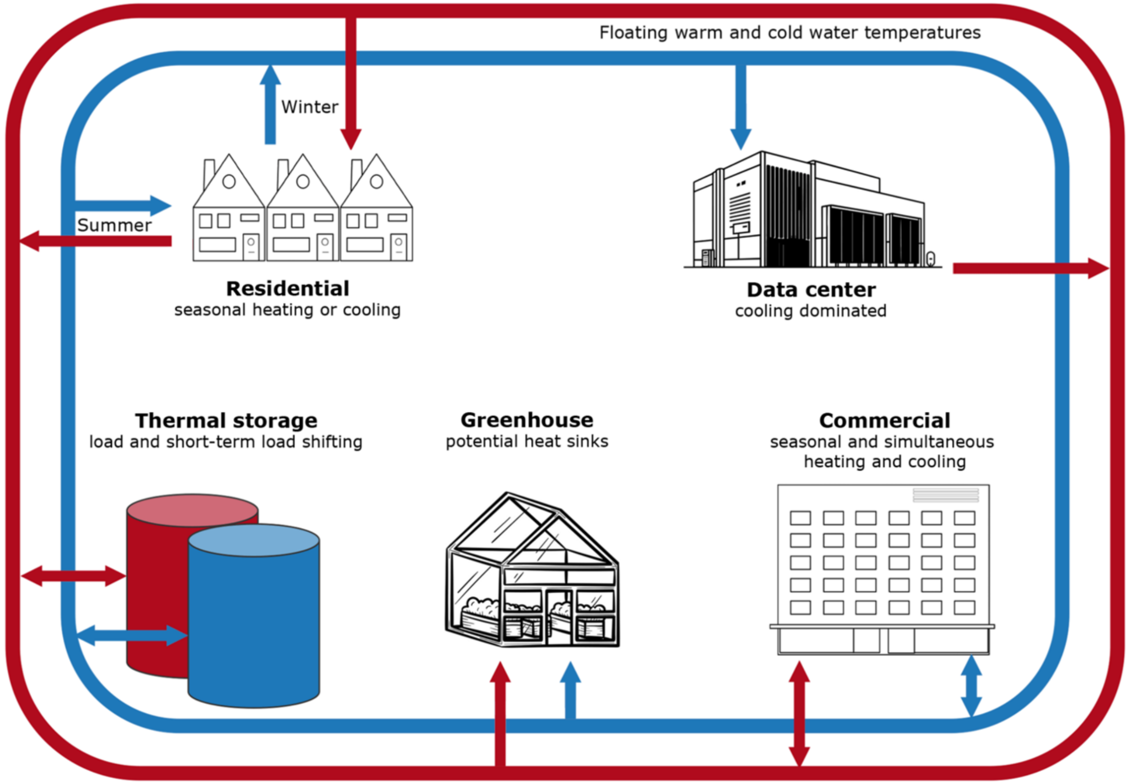
Data-centric scientific machine learning for urban intelligence and sustainability



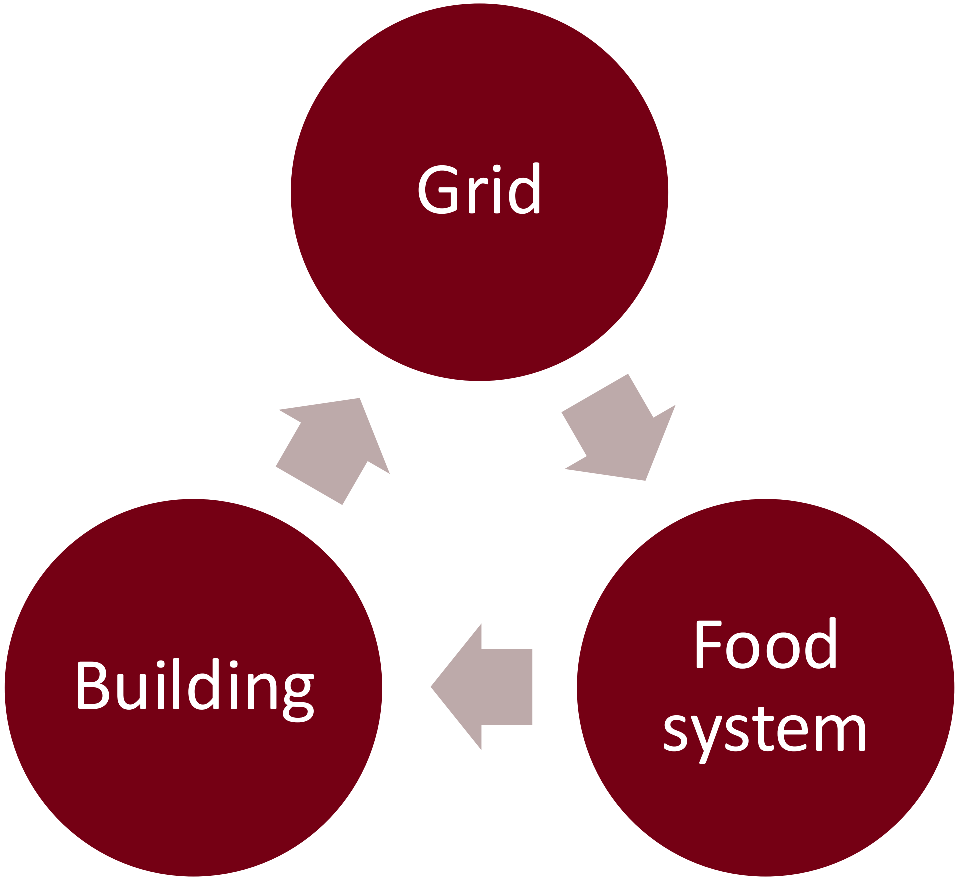
Synergetic urban farming and human environment



Data-centric scientific machine learning for urban intelligence and sustainability

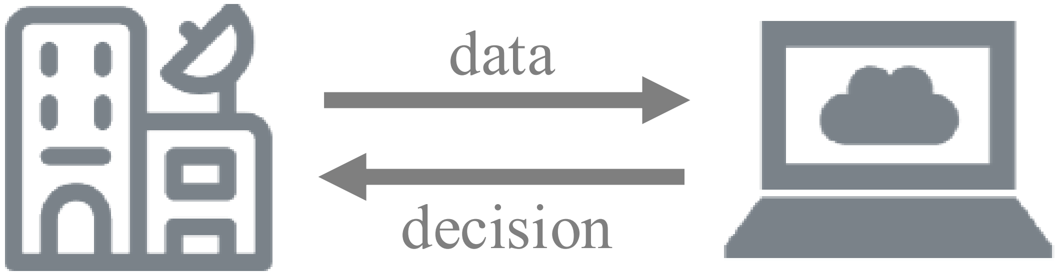


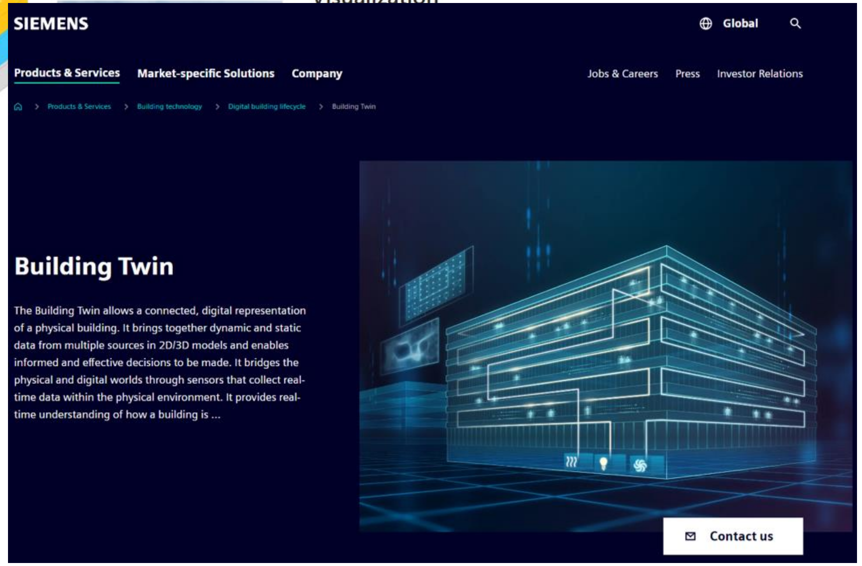
5GDHC, waste heat recovery and building electrification



Data-centric **scientific machine learning** for urban intelligence and sustainability

Digital twins - Computational models that replicate the behaviour of real-world systems, conducting virtual experiments in unseen scenarios and supporting decision-making

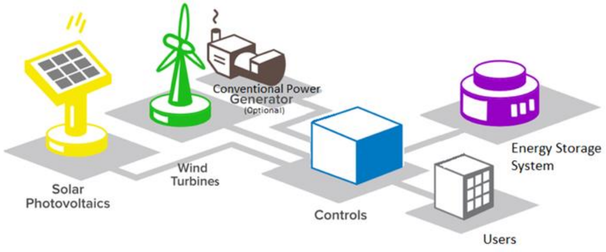




Data-centric **scientific machine learning** for urban intelligence and sustainability

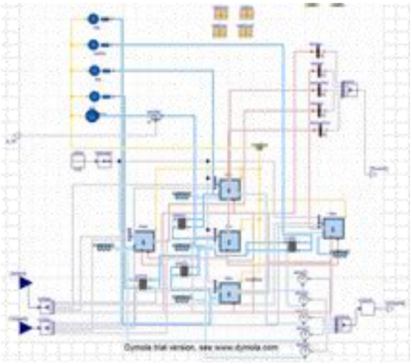
Informed decision-making requires scalable and extrapolatable predictive model

Open complex systems



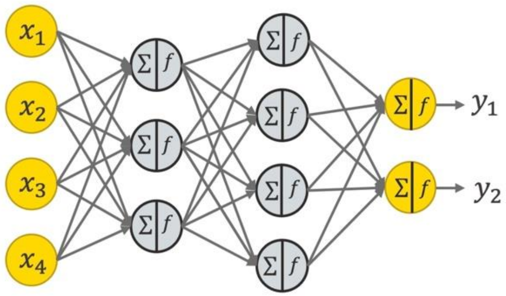
- Heterogeneous dynamics
- Subject to uncertainties
- Under-measured

Physics-based model



- Extrapolatable
- Low fitting capability
- Unscalable

Fully data-driven models



- High accuracy
- Scalable
- Unreliable in unseen scenarios

Data-centric **scientific machine learning** for urban intelligence and sustainability

Sci-ML via data assimilation:

- Extrapolation enabled by physics-informed constraints
- Expert-agnostics learning mechanism
- Balancing model complexity with data informativeness

Physics-informed models

Extrapolatable physics-informed models through data assimilation

+

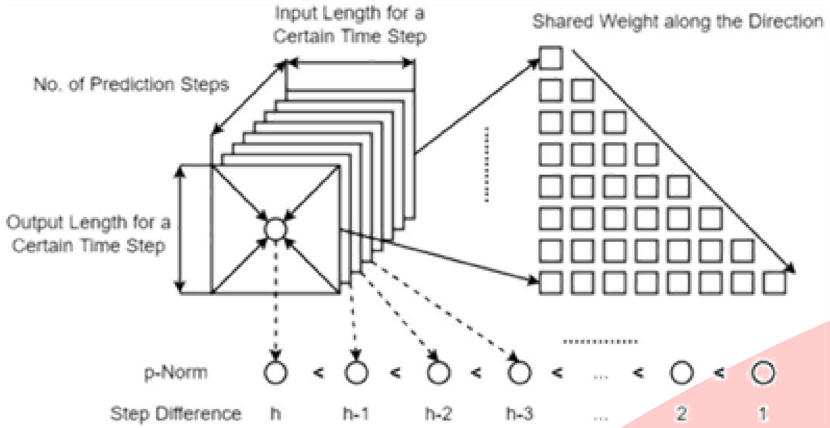
Decision-oriented data

Active-learned data acquisition to enable fit-for-purpose modeling

Dynamic digital twins

Dynamically evolving models to support decision-making throughout system life cycles

Data-centric **scientific machine learning** for urban intelligence and sustainability



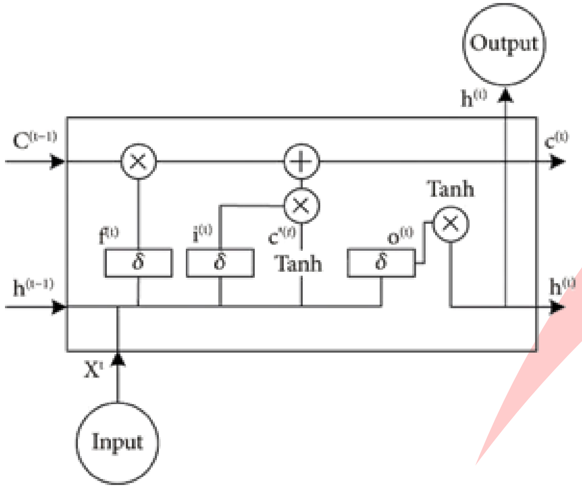
Increasing levels of physics

RC

PCNN

PINN

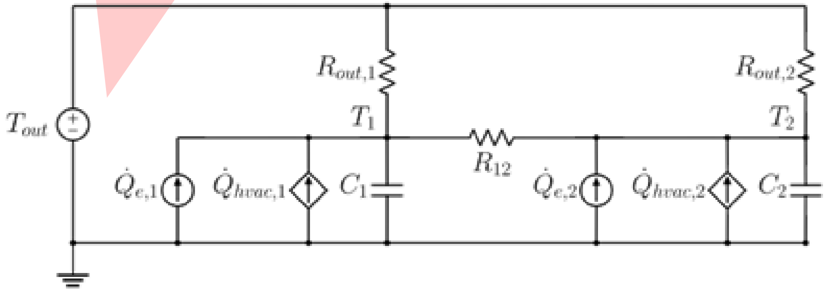
LSTM



$$\mathcal{L}_{PINN} = \mathcal{L}_{MSE} + \lambda \mathcal{L}_{grad},$$

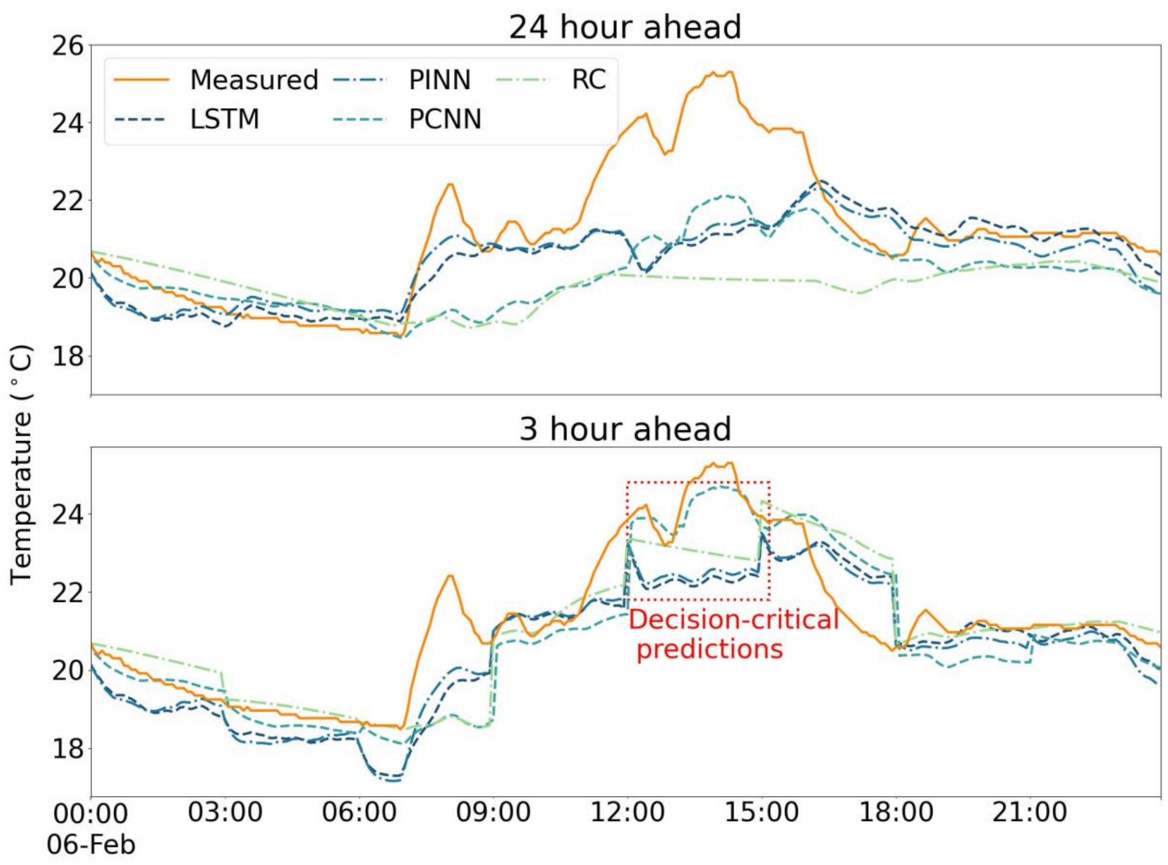
$$\mathcal{L}_{grad} = \frac{1}{l}\sum_{k=0}^{l-1}\left[\frac{1}{m}\sum_{z=1}^{m} g_k^z\right],$$

$$g_k^z = \sum_{y=1}^{m} ReLU\left(-\frac{\partial \hat{T}_l^z}{\partial u_k^y}\right) + ReLU\left(-\frac{\partial \hat{T}_l^z}{\partial T_k^{out}}\right)$$



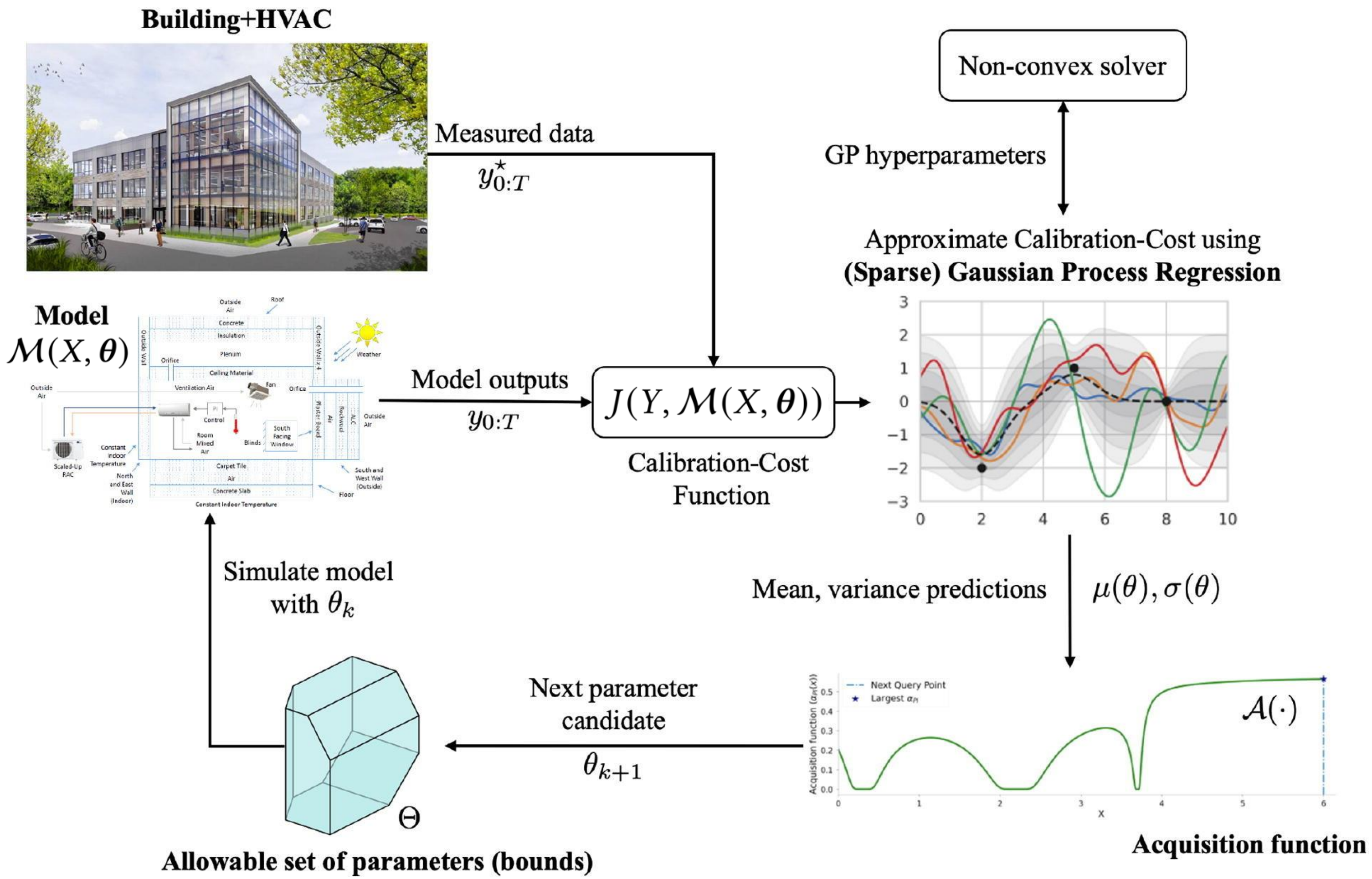
Data-centric **scientific machine learning** for urban intelligence and sustainability

	Traditional test		Extrapolation test	
	std	*avg*	*std*	*avg*
LSTM	0.21	0.89	0.44	1.21
PINN	0.23	0.89	0.51	1.12
PCNN	0.22	0.88	0.27	0.87
RC	0.15	1.39	0.13	1.78



Multi-zone thermal response model
- Essential to balance fitting capability and physics-based constraints
- Physics-consistency more important than predictive accuracy

Data-centric **scientific machine learning** for urban intelligence and sustainability



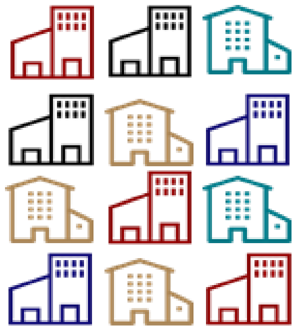
$$\theta^* = arg \min_{\theta \in \Theta} (J(\hat{Y}, Y))$$
$$s.t. \quad \hat{Y} = \mathcal{M}(X, \theta)$$

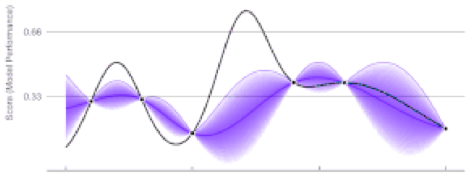
Define the optimization under limited information

- Develop base model
- Obtain X, Y
- Screen parameter

Data-centric **scientific machine learning** for urban intelligence and sustainability

Previous calibration runs:
source tasks





Calibration by Vanilla-Bayesian Optimization with Gaussian Process

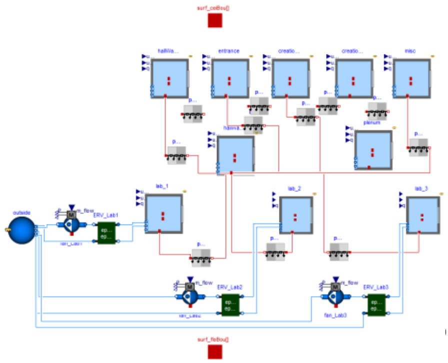
Training data storage

Source task data archived in the cloud: training set

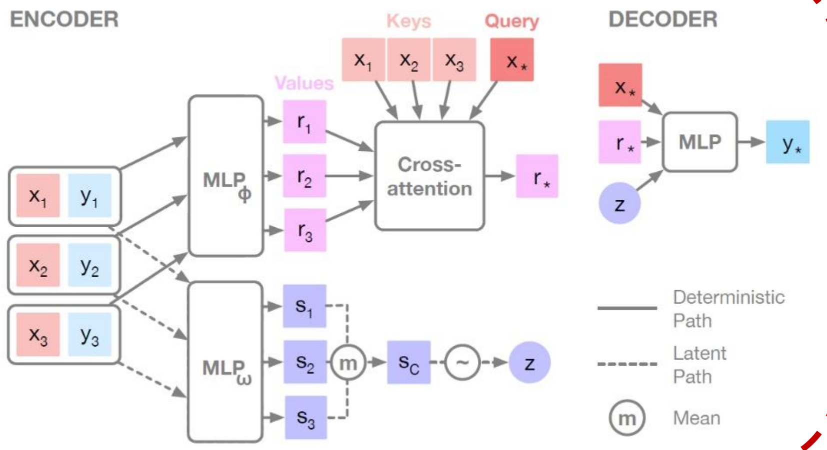
	θ_1	θ_2	...	θ_m	J
1	$\theta_{1,1}$	$\theta_{2,1}$	...	$\theta_{m,1}$	J_1
2	$\theta_{1,2}$	$\theta_{2,2}$	...	$\theta_{m,2}$	J_2
...		...			...
n	$\theta_{1,n}$	$\theta_{2,n}$	...	$\theta_{m,n}$	J_n

Meta-train

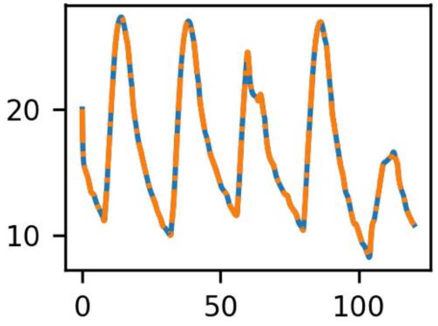
Previously unseen building data and model:
target tasks



Data-driven initialization



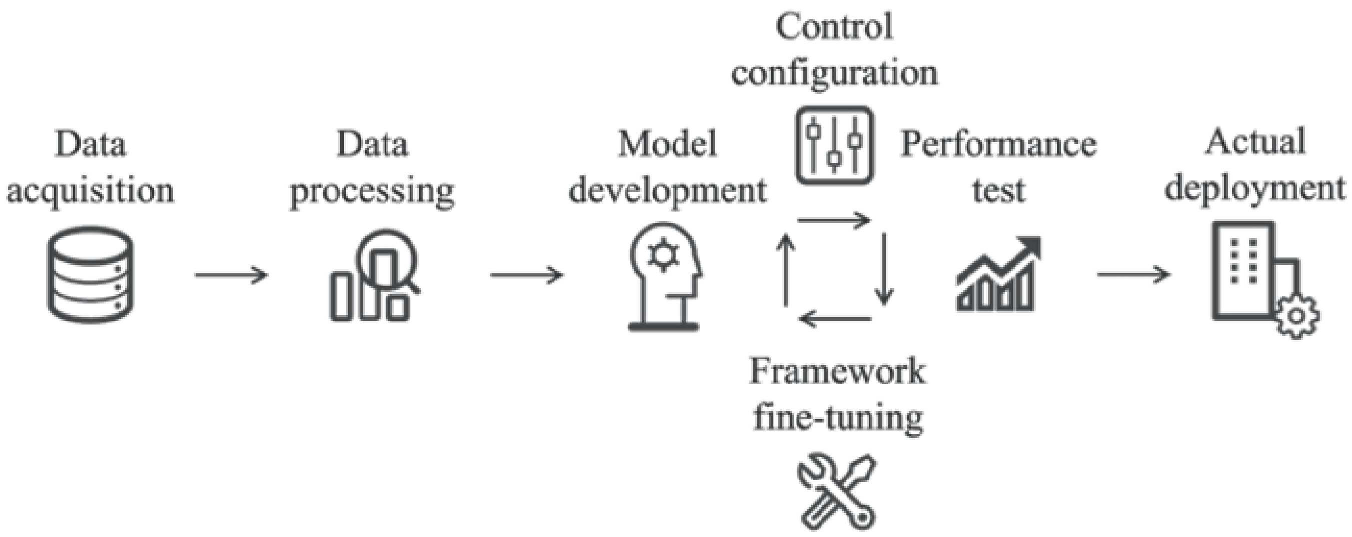
Meta-ANP-BO



Data-efficient learning

- Meta-learn the governing equations from a group of similar buildings
- Fastened target task with limited data
- 50% reduction in computational costs

Data-centric scientific machine learning for urban intelligence and sustainability

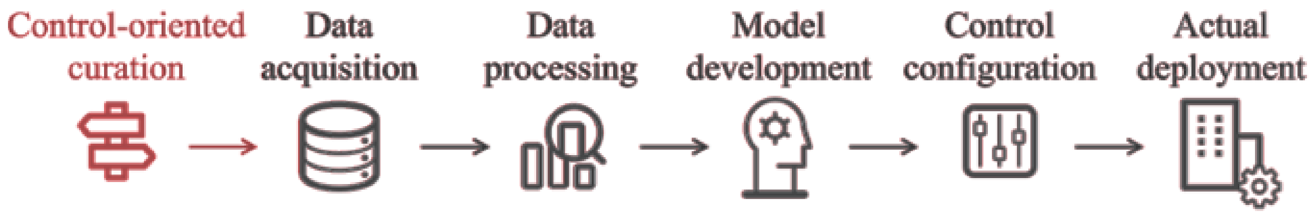


Traditional model-centric approach

- Arbitrary beforehand data acquisition
- Hard to reproduce
- Unknown consequence before deployment
- Low data usage rate

Data-centric scientific machine learning for urban intelligence and sustainability

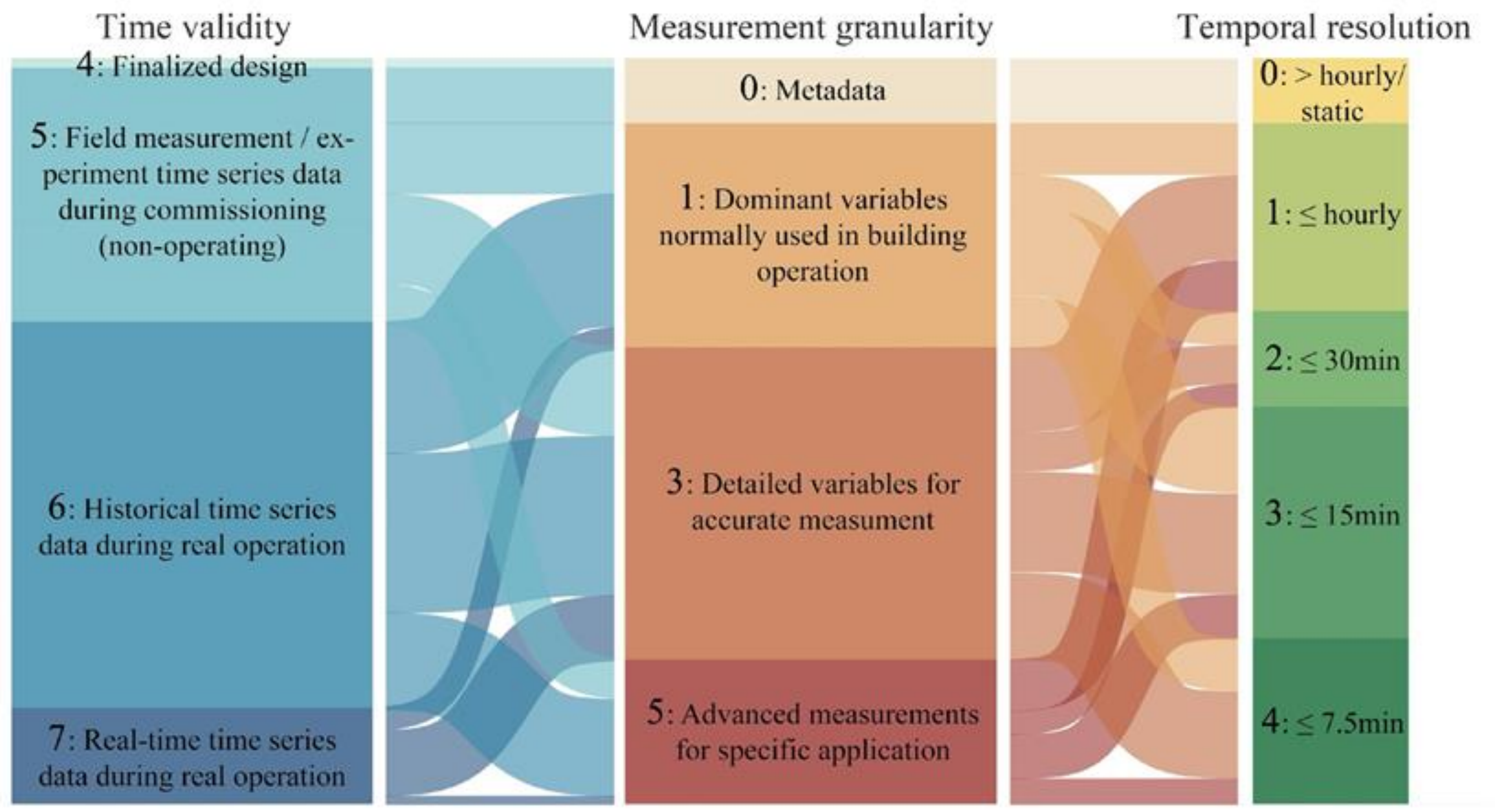
Data is like minerals, everywhere but few are useful . You need to know where to drill.



Data-centric approach

- Active learning guided by downstream applications
- Streamlined model development and implementation
- End-to-end performance estimation

Data-centric scientific machine learning for urban intelligence and sustainability



- Inherit from BIM LoD
- Three dimensions to define data availability
- Time validity: how up-to-date is the data
- Temporal resolution: time interval
- Measurement granularity: detailed definition for <u>6 data categories</u>
- Higher LoD corresponds to higher data acquisition cost

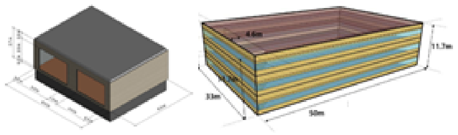
Data-centric scientific machine learning for urban intelligence and sustainability

Level	*0*	*1*	*3*	*5*[a]
Energy consumption	N/A	Total consumption by energy sources	Separated consumption by usage type	Separated consumption by sub-components
Indoor condition	N/A	Indoor air temperature	Variables affecting thermal comfort	Thermal comfort/ sensation feedback
System condition	System specifications	On/off operating mode, thermostat setpoints	Temperature and flow rate variables	Static pressures
Envelop condition	Geometric and thermal properties	N/A	Surface or core temperature	N/A
Internal disturbance	Assumed operating schedules/ profiles	N/A	Estimated operating profile	Additional occupant sensors
External disturbance	N/A	Weather data of the city/region	On-site weather station/sensors	Solar heat gain on different orientations

Data-centric scientific machine learning for urban intelligence and sustainability

Emulator setup

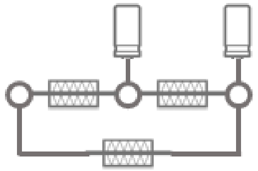
Measured disturbance inputs
- Actual meteorology year
- Occupant and electricity (office and classroom)



High-fidelity models
- BESTEST case600
- DOE medium office prototype

Evaluated against the emulators

Model identification

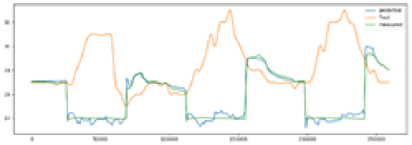


$$\theta = argmin \int_{t0}^{t1} \sum_{i}^{k} (T_{room,i} - \hat{T}_{room,i})^2$$

Control-oriented models
- RC model complexities
- Inputs for internal heat gain
- Open-loop prediction error

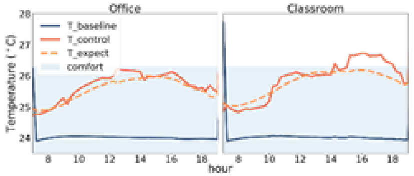
Evaluated against the emulators

Predictive evaluation



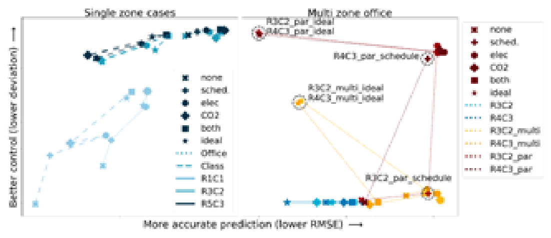
- Root mean square error
- Tested under multiple operating conditions

Control evaluation



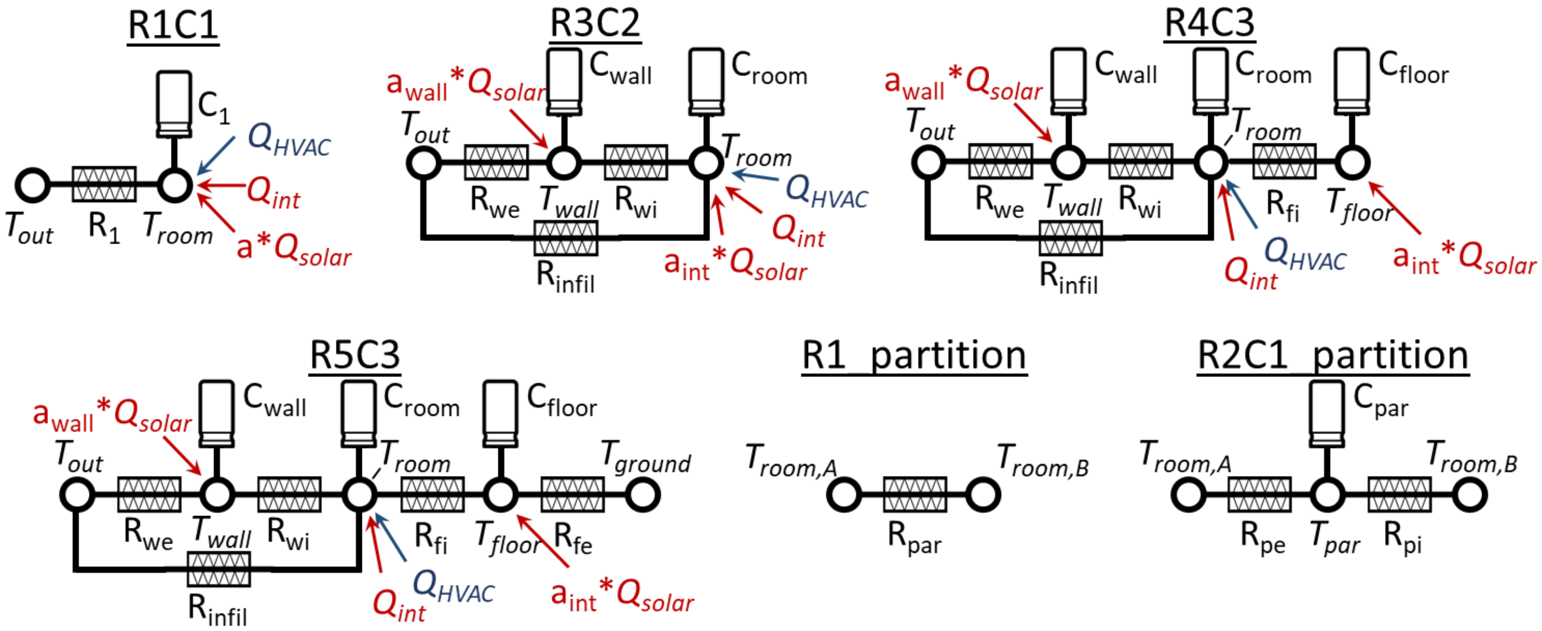
- Minimize overall penalty of energy and comfort
- Track multiple setpoints

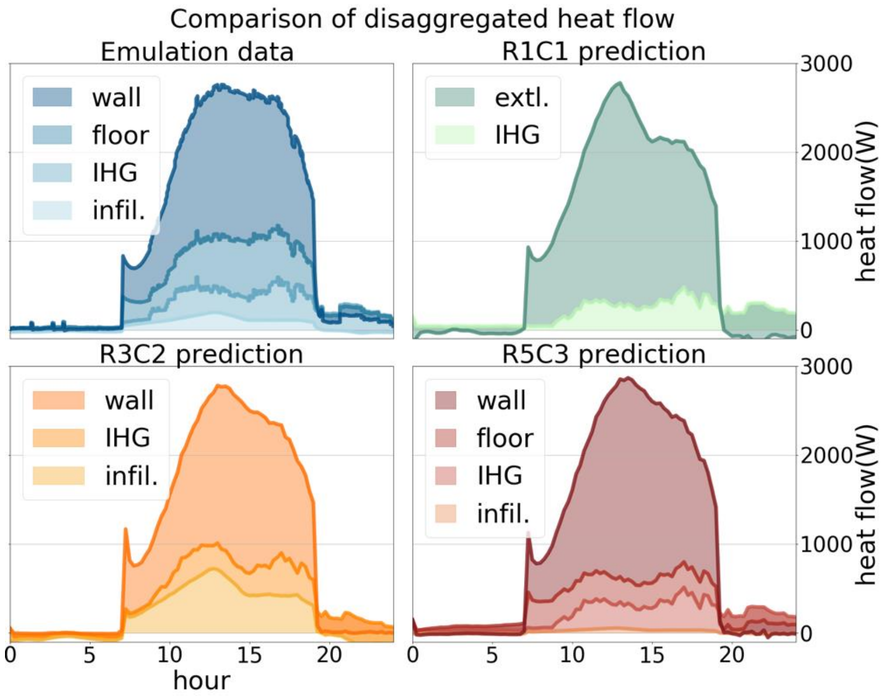
Post analysis



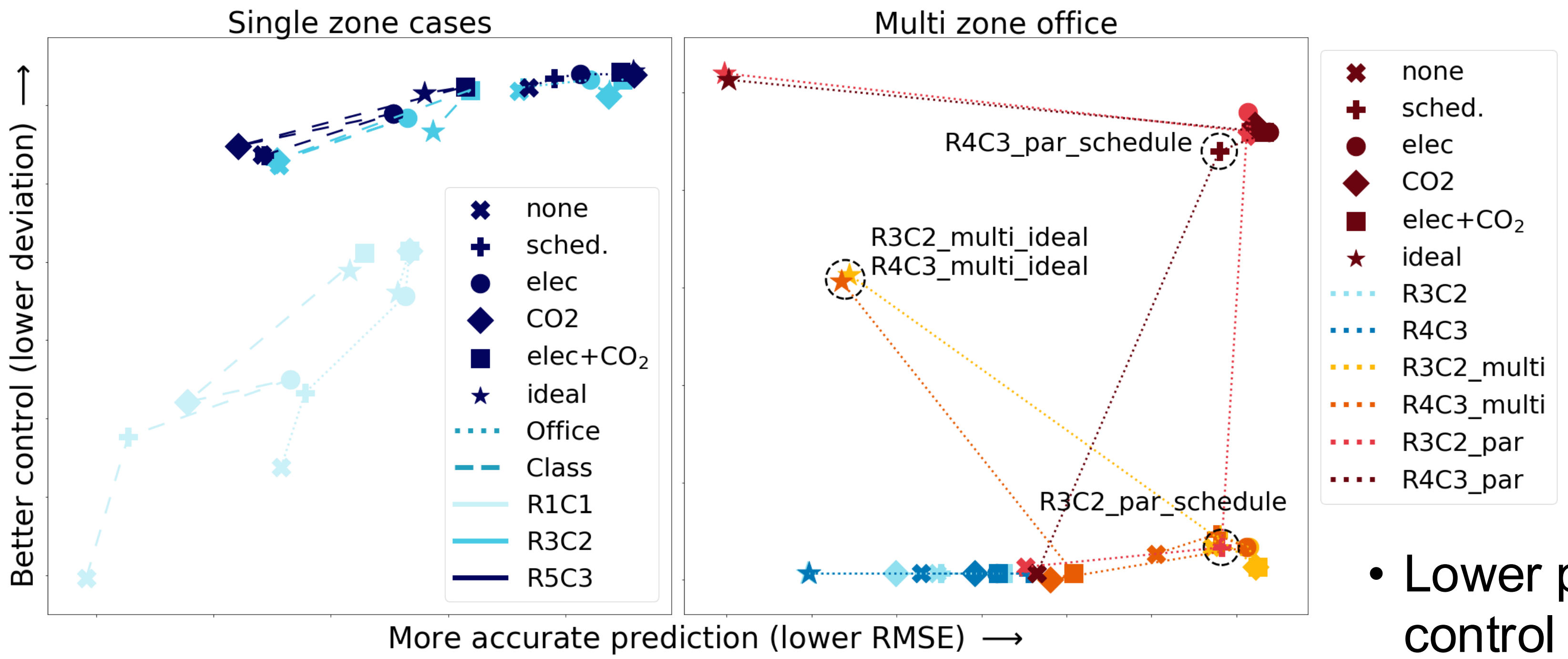
- Affecting factors of data usefulness
- The interaction between data and model in identification
- Heterogeneous requirements of prediction and control

Data-centric scientific machine learning for urban intelligence and sustainability



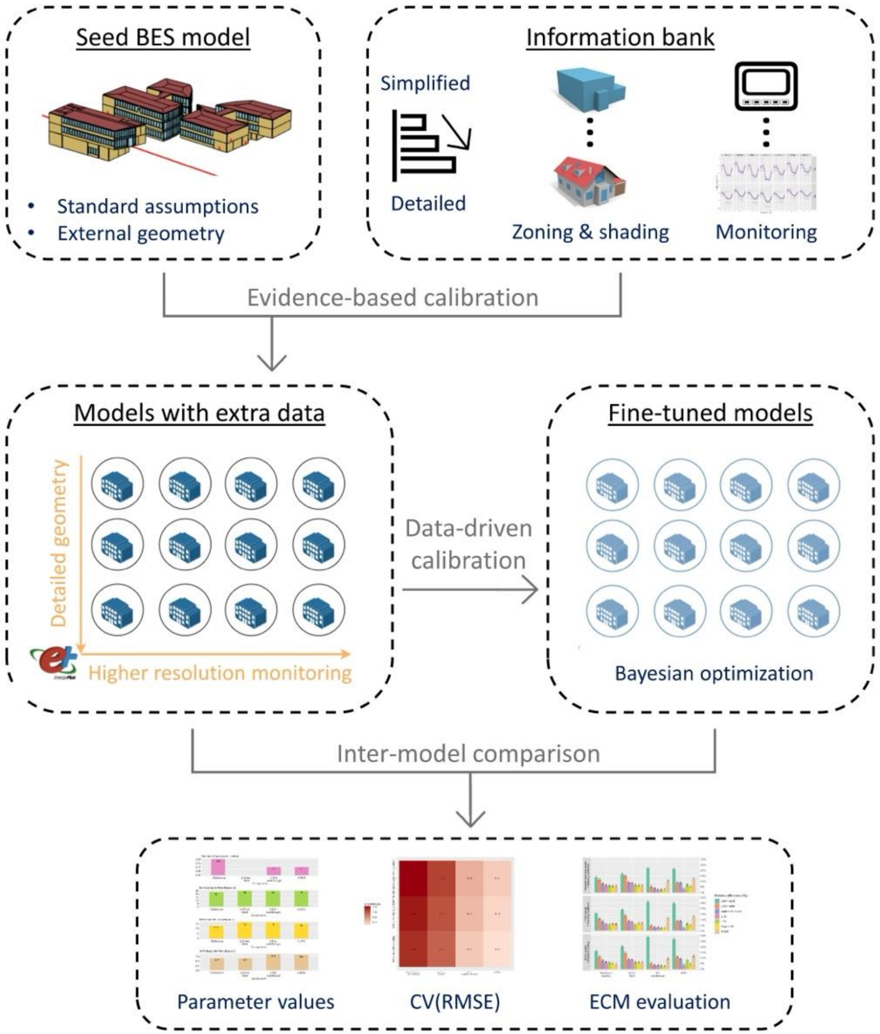


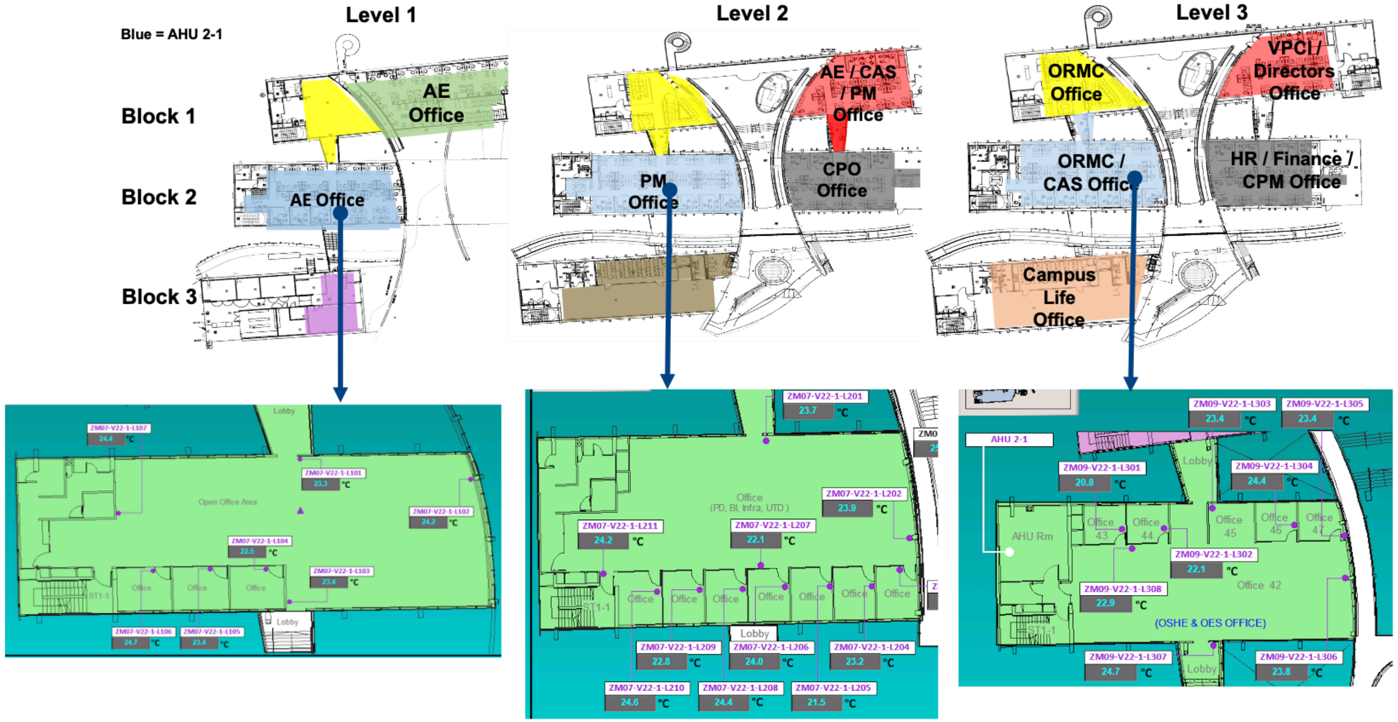
Data-centric scientific machine learning for urban intelligence and sustainability



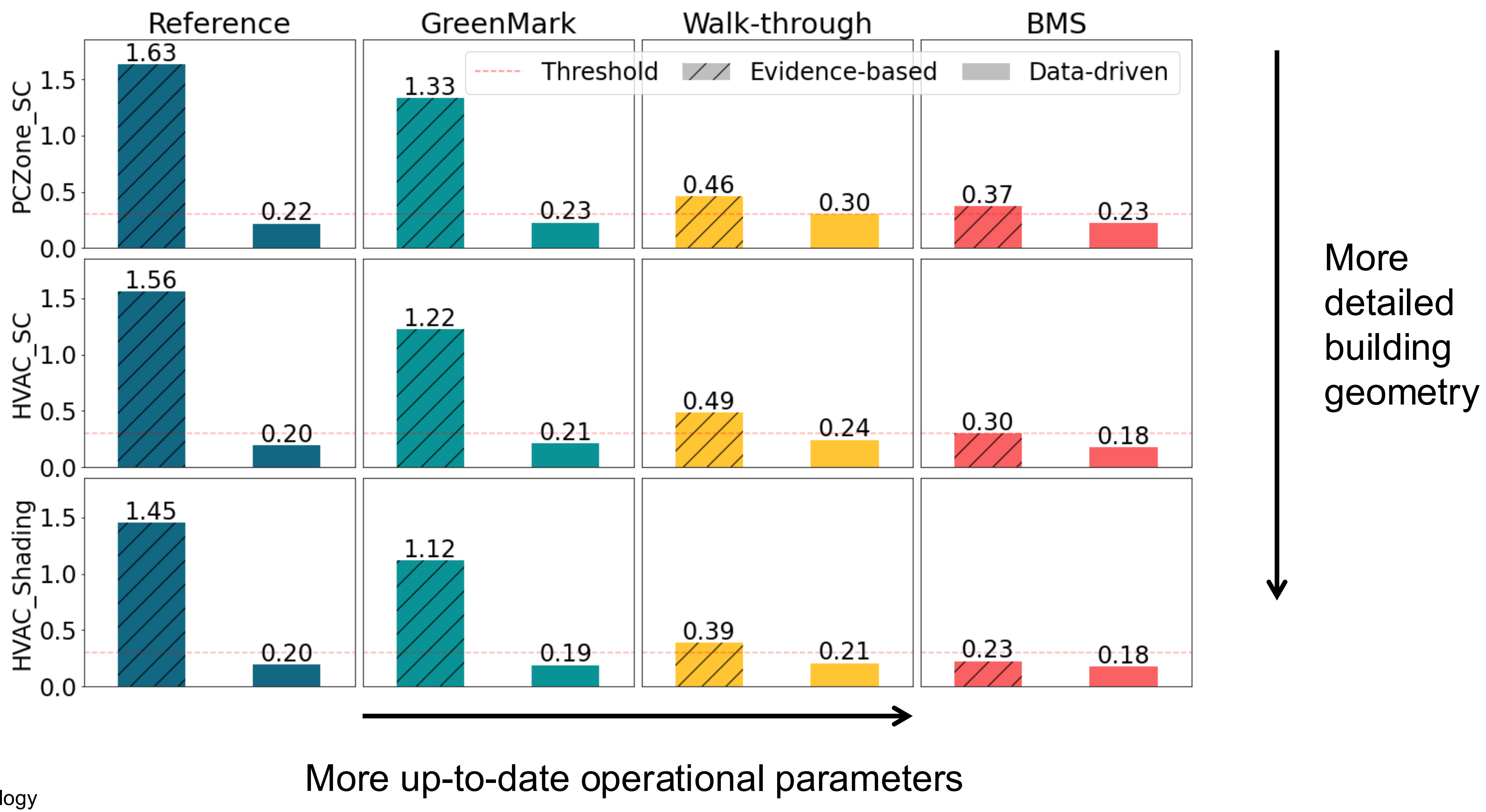
- Lower prediction error means better control for simple dynamics
- For complex buildings, more data only led to lower predictive errors
- Better control requires adequate model
- Critical physical component should be preserved (partition capacitor here)

Data-centric scientific machine learning for urban intelligence and sustainability

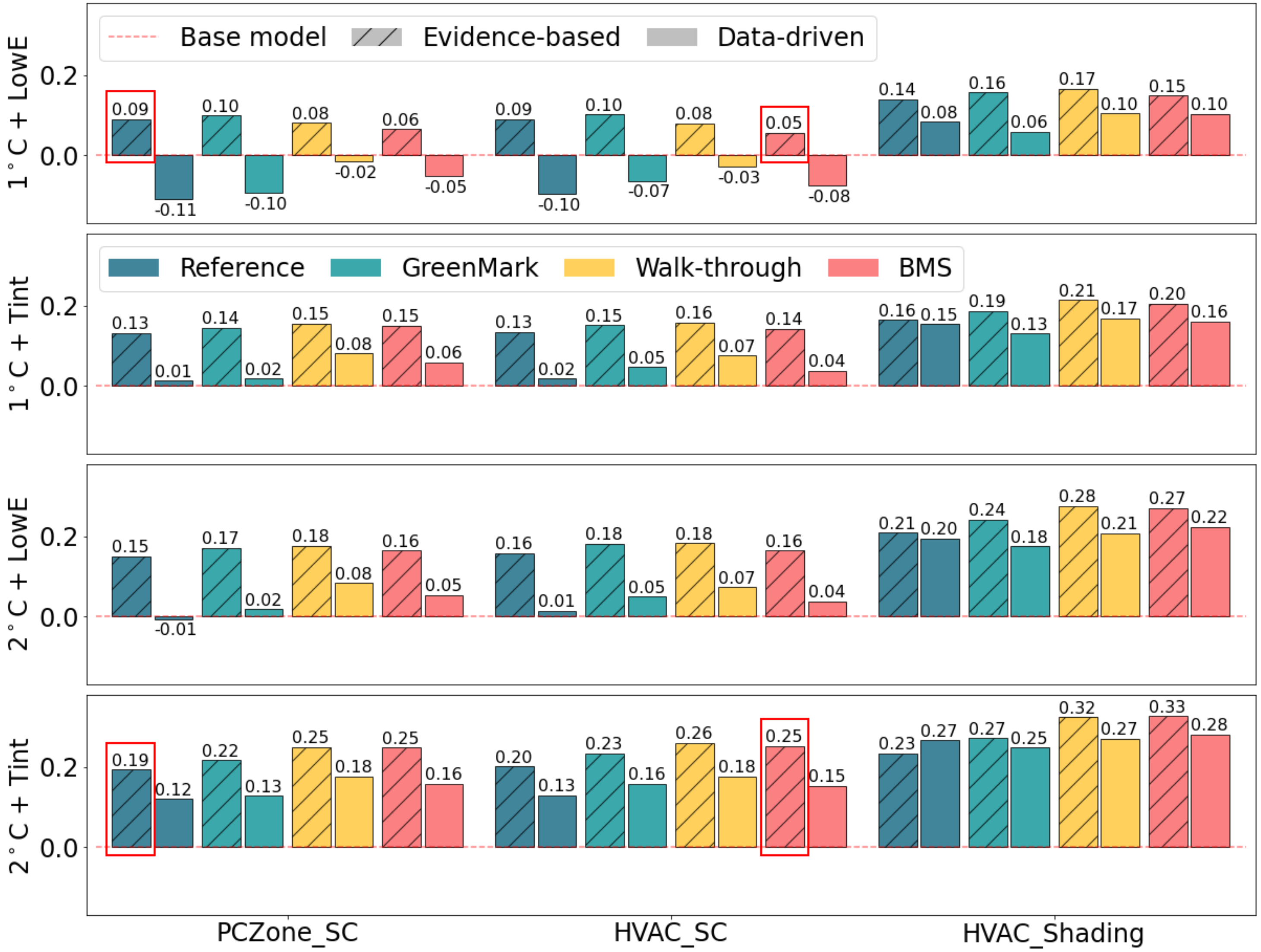




Data-centric scientific machine learning for urban intelligence and sustainability



Data-centric scientific machine learning for urban intelligence and sustainability



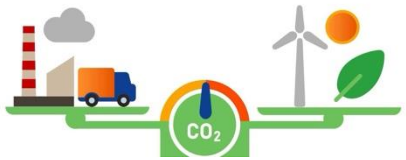
- Models with similar accuracy could have distinct ECM evaluation
- The key to informed decision-making is a good estimate of the relevant parameter

Data-centric scientific machine learning for urban intelligence and sustainability

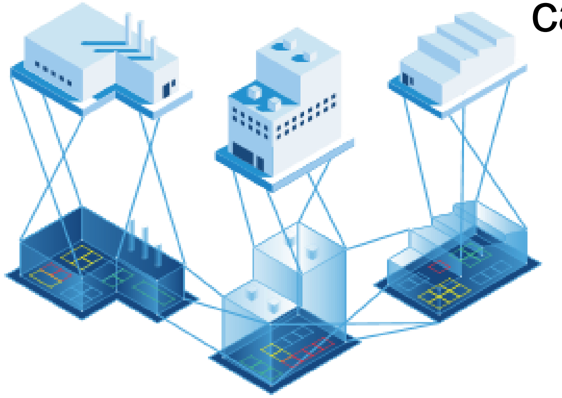
- Be cautious about low predictive errors in digital twin applications
- Scalable DT frameworks should start from data exchange infrastructure
- Model development needs to be fit-for-purpose
- Multi-system and cross-scale applications remain challenging

Thank you!

Human societal activities

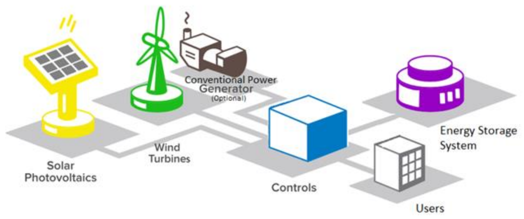


Energy-food-carbon nexus



Cross-scale digital twins for urban sustainability

Carbon-aware system operations



Optimal dispatch of hybrid energy

Sicheng Zhan

szhan@mit.edu

Massachusetts Institute of Technology