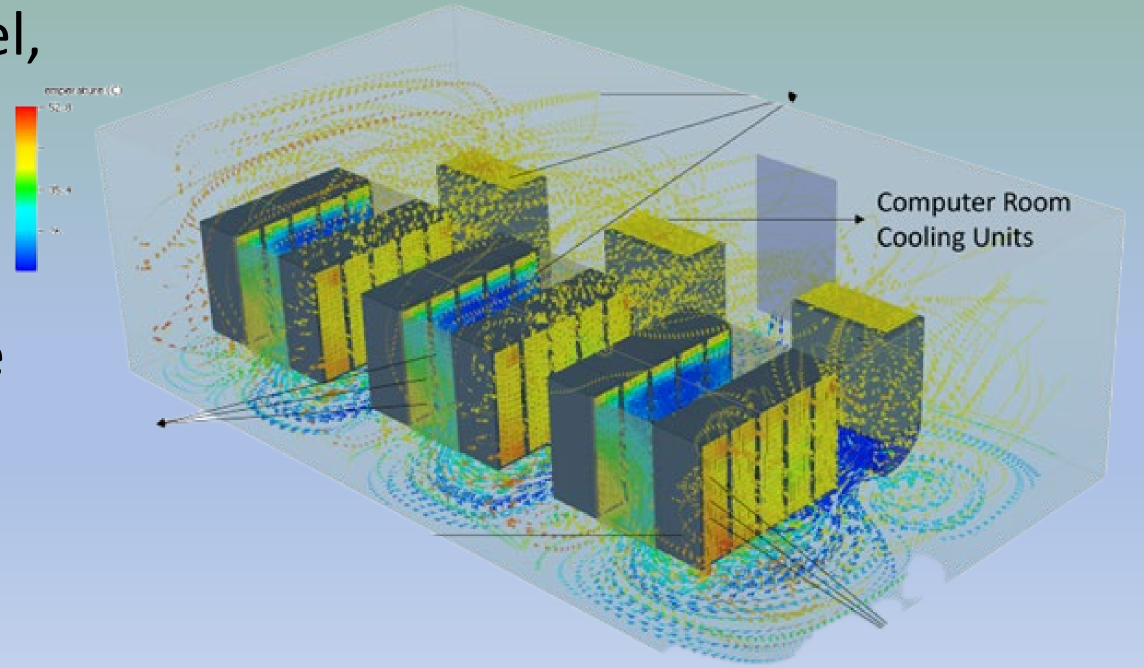
Enhancing Data Center Sustainability with a 3D CNN-Based CFD Surrogate Model

NEURAL INFORMATION PROCESSING SYSTEMS

Climate Change AI

Soumyendu Sarkar, Avisek Naug, Zachariah Carmichael,
Vineet Gundecha, Ashwin Ramesh Babu,
Antonio Guillen, Ricardo Luna Gutierrez

Hewlett Packard Labs @ Hewlett Packard Enterprise



Hewlett Packard
Enterprise

Motivation

Models Data Center Thermodynamics

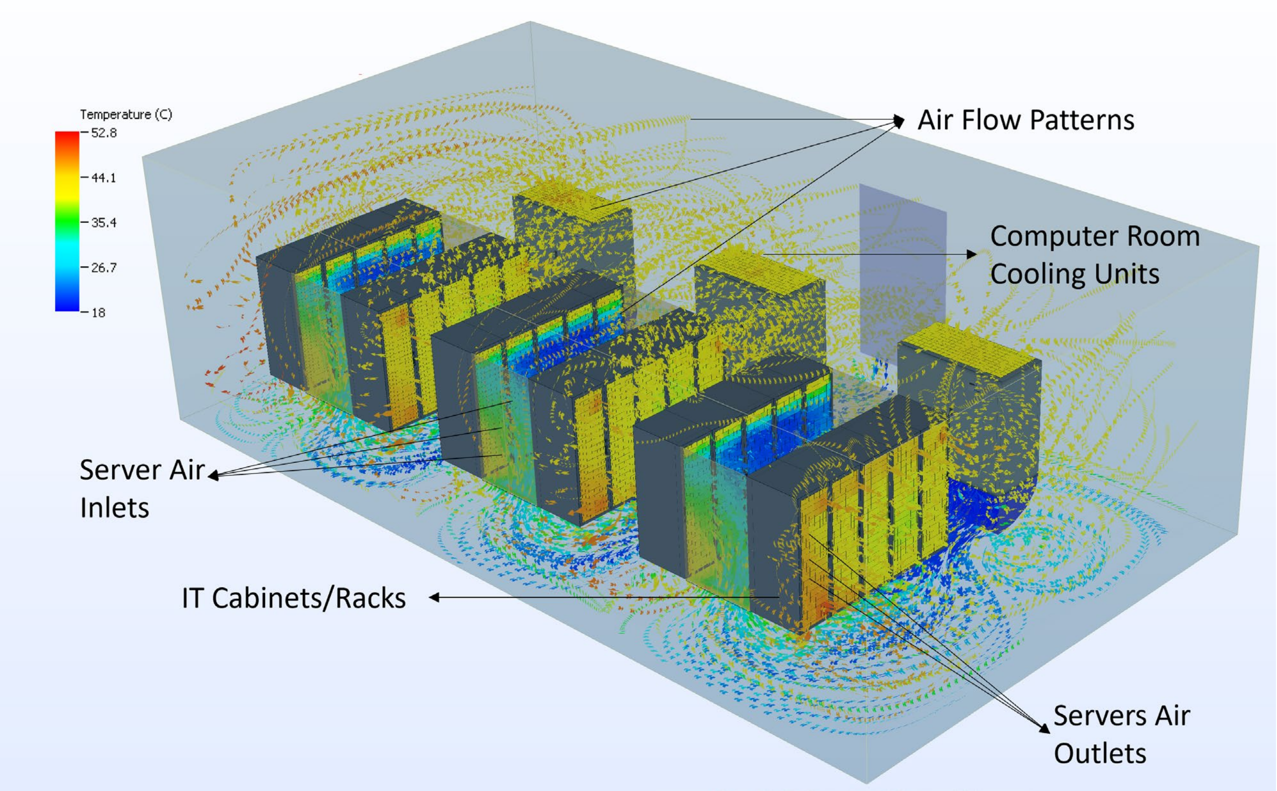
Reveals areas where equipment is overheating or underutilizing cooling resources

Optimize data center cabinet and server placement,

Prevent hot-spot-related server

Increase equipment longevity.

A **fast thermal analysis can enable control, which will enable real-time control to significantly lower carbon footprint.**

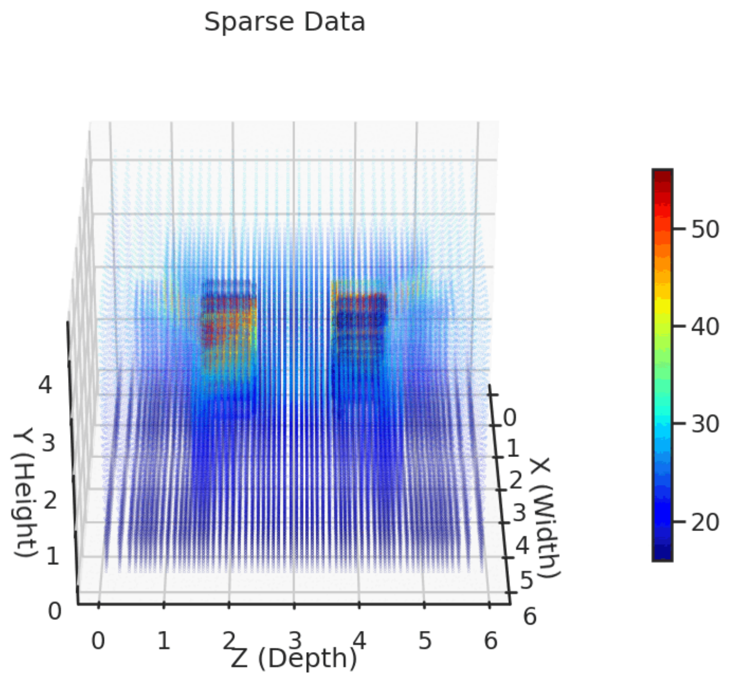


Data Center Heat Flow

Challenges of Traditional CFD and Possibilities

What is the Challenge?

- ❑ High computational complexity
- ❑ Not scalable for real-time applications
- ❑ Fast ML surrogates make real-time control possible
- ❑ ML surrogate can scale to very large data centers



Generated by ML Surrogate for CFD

3D CNN Surrogate: Input Configuration

3D Input Data:

Power: volume-normalized power

ACU Set Point: Encodes the normalized ACU temperature

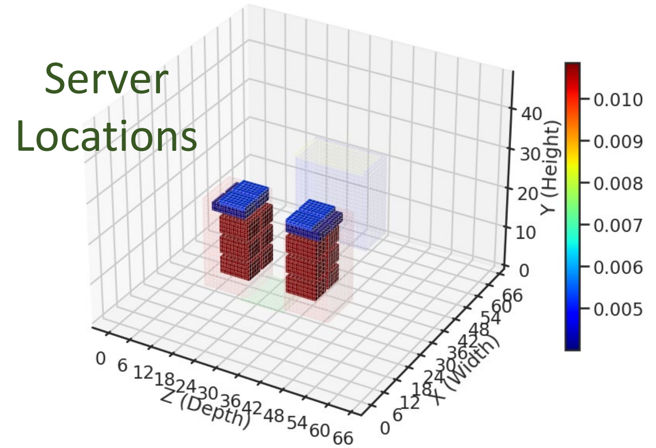
ACU Fan Speed: Volume-normalized fan

3D Output Data:

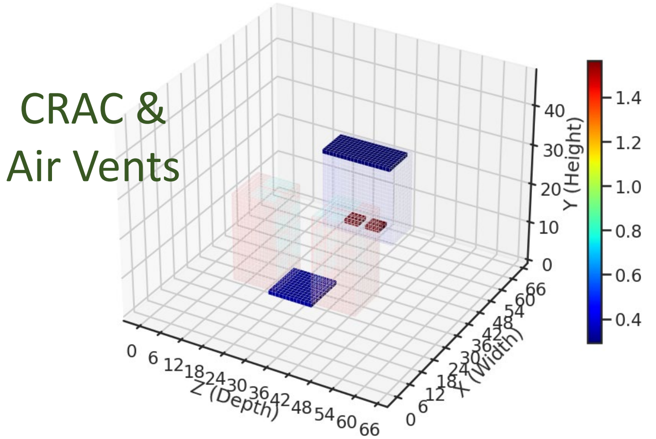
3D Heat Map

3D CNN Surrogate of CFD

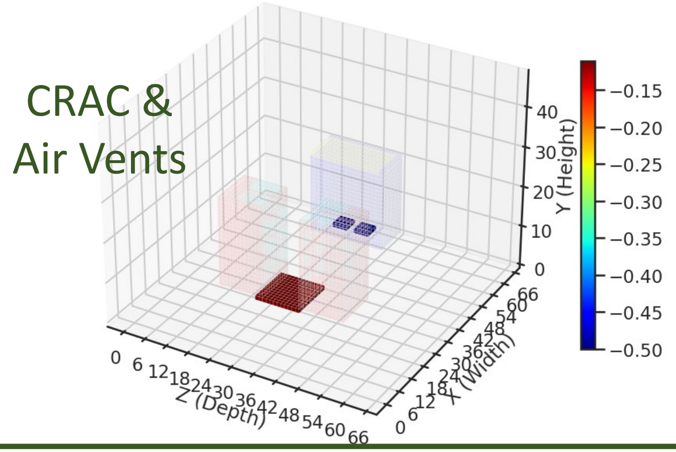
Channel #1: Heat/Power



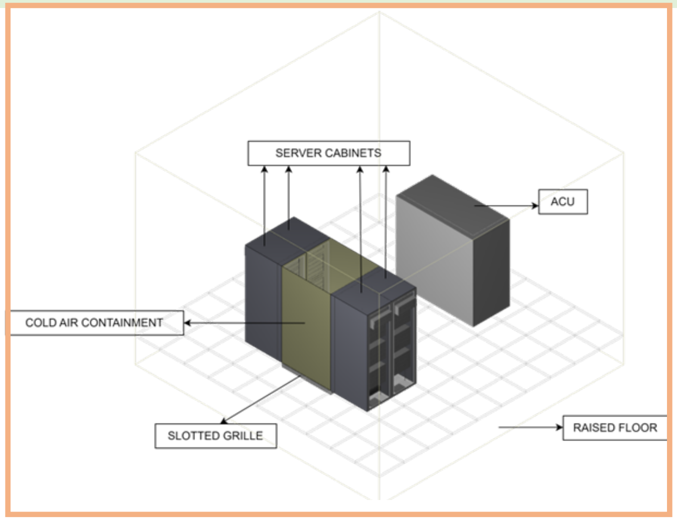
Channel #2: Fan Speed



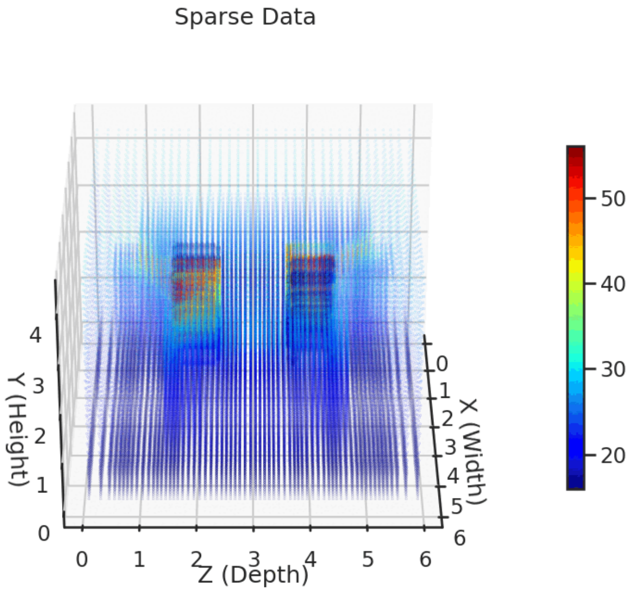
Channel #3: HVAC Set Point

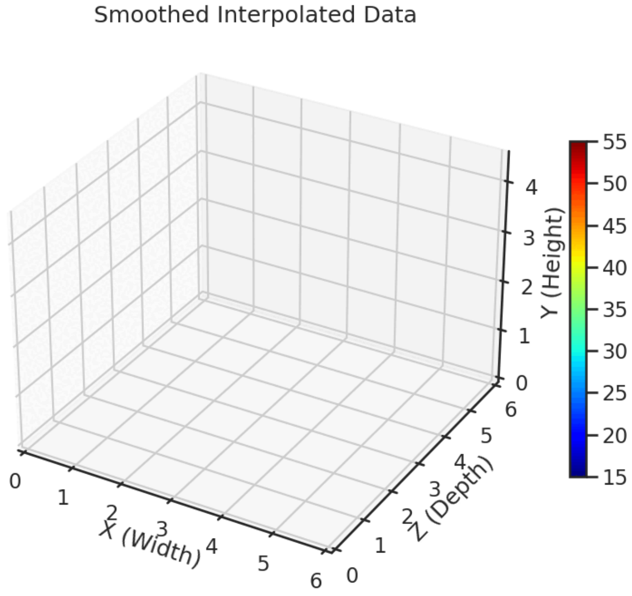


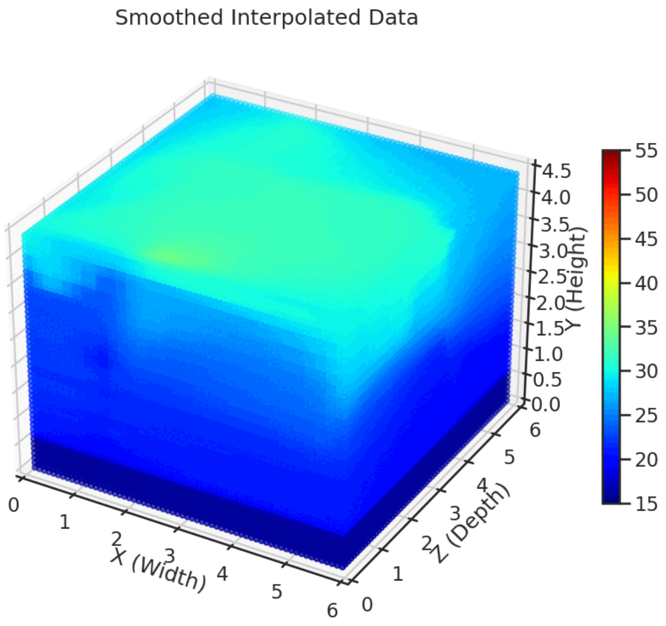
3D CNN Surrogate: Output Generation



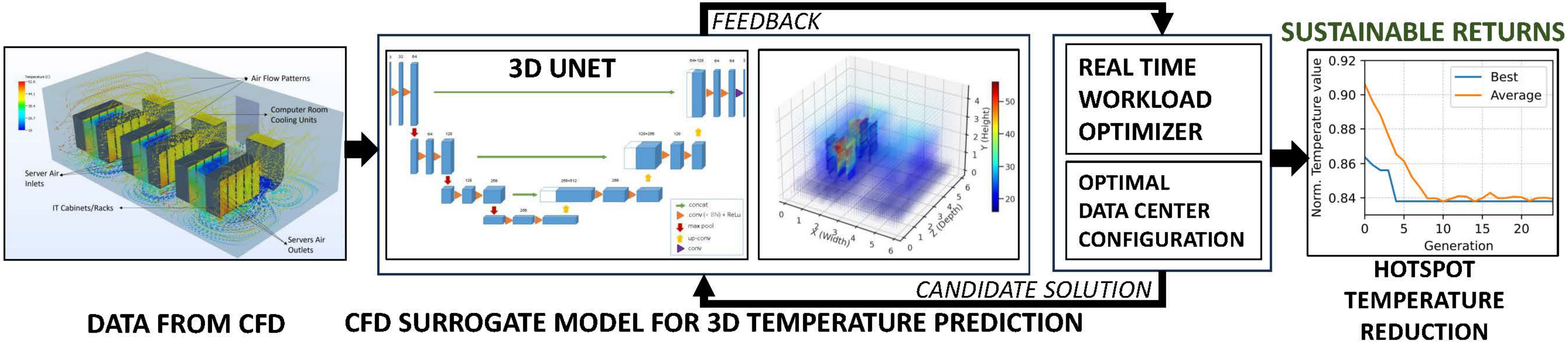
Data Center Configuration







Sustainability Advantage



Model Accuracy and Genetic Algorithm of the Surrogate Model

Model accuracy evaluation metrics : To evaluate the effectiveness of our approach, we consider several metrics: inference time, mean-squared-error (MSE), mean maximum absolute error (AE), mean AE, top-t AE, and 3D structural similarity (SSIM). Top-t AE is especially useful for evaluating how well our approach models hot spots in data centers. It is defined below:

$$\text{Top-}t\ \text{AE}(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) = \sum_{n=1}^{N} \sum_{i,j,k \in \text{top-}t(\boldsymbol{y}_n)} \frac{|y_{n1ijk} - \hat{y}_{n1ijk}|}{Nt} \tag{1}$$

The function top-$t(\cdot)$ returns the 3D indices of the largest t elements of its tensor argument, i.e., the hottest t temperature locations of a ground truth heatmap. In our evaluation, we set t to 10% of the number of heatmap elements. The inference time is the time to predict a single sample after the model and data are loaded into GPU memory.

- Goal: Optimal workload distribution → Reduce Temperature Hotspots
- Genetic Algorithm Optimizer
- Objective Function : Trained surrogate model
- Reducing hotspots by approximately 7.70% within 25 iterations .
- 99% reduction in optimization time

Evaluation of Speed and Accuracy of the 3D CNN Surrogate Model

Workload Settings	Model	Inf. Time (ms)	MSE	Mean AE (C)	Max AE (C)	Top-t AE (C)	3D SSIM
Uniform Utilization	Res. U-Net	41.6	**0.00086**	**0.87**	**9.99**	**1.29**	**0.9522**
	U-Net	40.6	0.00094	0.93	11.0	1.34	0.9437
	V-Net	35.6	0.00112	1.44	12.8	2.32	0.9112
	SegNet	**15.2**	0.00432	2.06	14.7	6.27	0.8434
Extreme Utilization	Res. U-Net	41.6	**0.00010**	**0.34**	**4.65**	**0.37**	**0.9772**
	U-Net	40.6	0.00031	0.57	9.90	0.62	0.9565
	V-Net	35.6	0.00048	0.72	6.66	0.79	0.9343
	SegNet	**15.2**	0.00137	1.25	8.83	2.74	0.9028
Grid CPU Utilization	Res. U-Net	41.6	0.00118	0.84	8.90	1.68	0.9459
	U-Net	40.6	**0.00018**	**0.40**	**7.32**	**0.53**	**0.9798**
	V-Net	35.6	0.00312	1.93	14.5	2.35	0.8576
	SegNet	**15.2**	0.00559	2.27	19.7	5.26	0.8197
Grid CPU Utilization and Cold Aisle Containment	Res. U-Net	41.6	0.00011	**0.21**	**4.47**	**0.43**	**0.9939**
	U-Net	40.6	**0.00009**	0.23	6.39	0.44	0.9911
	V-Net	35.6	0.00176	1.30	9.40	1.84	0.8962
	SegNet	**15.2**	0.00315	1.11	20.7	4.73	0.9220

Table 1: Evaluation (Speed and Accuracy) metrics on test set. The inference time is computed on a V100 GPU.

2D slices of 3D CNN Surrogate Model Output

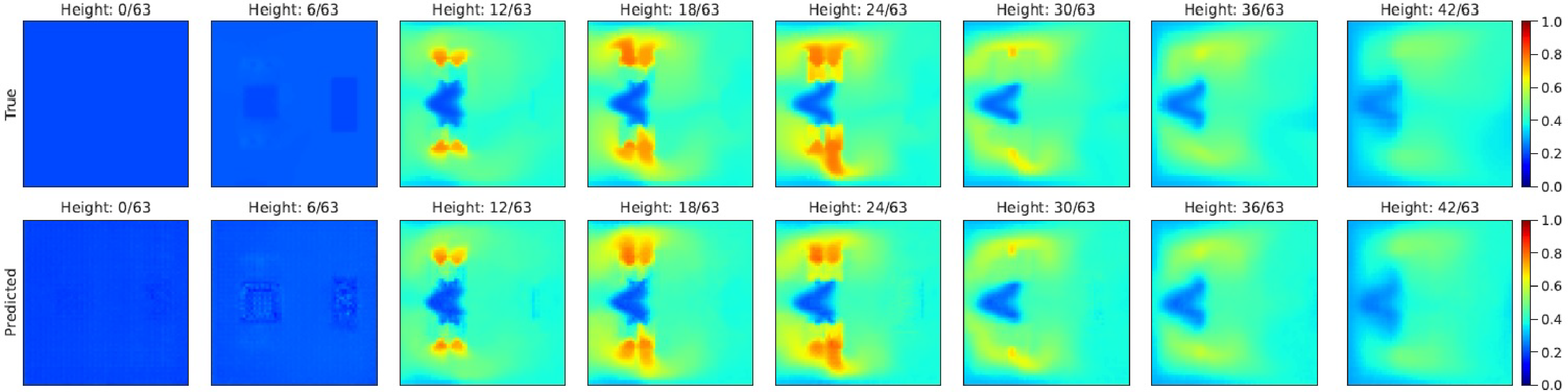


Figure 3: Matrix representation across different slices of the data center room. True outputs are in the top row, and model predictions the bottom row. Each column represents a slice at varying heights.

Impact of CFD Surrogates for Data Centers on Sustainability

Applying our optimized workload configurations derived from 3D CNN and the genetic algorithm, we got:

- significant improvement in sustainability and energy efficiency
- with an optimal workload distribution, the HVAC cooling **energy consumption rates drops,** translating to cost savings and a notable carbon footprint reduction.
- the enhanced temperature regulation may lead to prolonged equipment lifespans and reduced e-waste and manufacturing carbon footprint, emphasizing the holistic benefits of our approach.
- the dual benefits of **financial savings** and **environmental impact** through optimized configurations
- as a future work, an expanded CFD surrogate will enable thermal analysis of large data centers

Thank You

Soumyendu Sarkar
Hewlett Packard Labs @ Hewlett Packard Enterprise