ShadingZip: Within-Building Selective Computation of Shading Profiles for Urban Building Energy Models

Chengxuan Li[1,*], Zihan Jimmy Wang[1], Timur Dogan[1]

[1]Environmental Systems Lab, College of Architecture, Art and Planning, Cornell University

ABSTRACT
Urban building energy models require time-resolved solar access at the envelope sensor level, making exhaustive ray tracing expensive. ShadingZip reduces ray-tracing evaluations by 67.4% at a 2.6% WMAPE by selectively computing shading for representative sensors within each building. Representatives are identified using sensor metadata (MF), exploratory shading observations (SF), or both (MF+SF); unobserved periods are reconstructed from assigned representatives. ShadingZip is evaluated on 2,691 buildings from ten morphological contexts of over 65,000 buildings under five latitudes. Accuracy is measured using error at building- and surface-aggregation levels. SF and MF+SF outperform MF once the cost of exploratory ray tracing is accepted, although MF+SF provides no consistent advantage over SF. A balanced configuration—SF with 3 exploratory periods and representative ratio 0.20—used 32.6% of exhaustive ray-tracing, with WMAPE below 5% in 49 of the 50 case–latitude combinations. ShadingZip provides a scalable and controllable tradeoff between shading accuracy and workload while preserving full building and surface shading result coverage.

KEYWORDS
Urban Building Energy Modeling, Solar shading, Clustering, Selective computation

INTRODUCTION
Urban building energy models (UBEMs) extend building physics simulation to urban scales, supporting retrofit prioritization and planning of distributed renewable energy resources (Hong et al. 2020). In UBEMs, solar exposure is an important boundary condition: radiation gains enter heat balance models and drive Building-Integrated Photovoltaic (BIPV) yields. This exposure depends on obstruction of buildings, terrain, and vegetation, which is costly to resolve as each envelope sensor requires repeated visibility tests. Fast pixel-counting, ray-tracing, and voxel approaches reduce shading computation cost (Hoover and Dogan 2017; Luo et al. 2020; Su and Dogan 2026). These approaches reduce the cost of an individual visibility evaluation, but urban-scale workloads still multiply that evaluation across many sensor-period combinations. Dimensionality reduction has been adopted to identify representative daylight distributions (Kent et al. 2020), yet this post-hoc approach does not reduce the initial simulation workload, motivating an a priori method that selects representative sensor-period combinations before the remaining shading calculations are performed. We therefore introduce ShadingZip

* Corresponding author email: cl2749@cornell.edu

to reduce shading calculation effort by selectively computing sensors. Sensors are grouped using metadata or exploratory shading results for limited periods. Full-year computations are only performed for the representative sensors whose profiles are used to reconstruct other sensors within each building. Features, the number of exploratory periods, and the representative ratio are varied to explore the tradeoff between avoidable ray-tracing cost and reconstruction accuracy.

METHODOLOGY

ShadingZip (Figure 1) operates on sensor sunlit fraction profiles. Given envelope sensors $\mathcal{S}_b$ of building b, for each sensor $i \in \mathcal{S}_b$, modeled period $t \in \mathcal{T}$, and hour h, the shading calculation produces sunlit fraction $y_{i,t,h} \in [0,1]$. Each profile contains 19 periods with 24 hourly values. Diffuse irradiance and sky view factors are neither calculated nor reconstructed. We compute complete profiles as ground truth to evaluate ShadingZip, which selectively computes a subset of sensor-periods, driven by clustering over feature family F (***Table 1***), with m exploratory periods $\mathcal{E}$, at representative ratio r to derive representative sensors from $\mathcal{S}_b$; F may include Metadata Features (MF), Shading Features (SF), and their combination (MF+SF). MF requires no ray-tracing. SF and MF+SF use $m \in \{1,2,3\}$, emulating summer solstice, both solstices, and equinox-and-solstices exploration periods. The process is detailed in Algorithm 1 and returns the representatives $\mathcal{R}_b$, assignments a, and observations $Y_{\mathcal{E}}$.

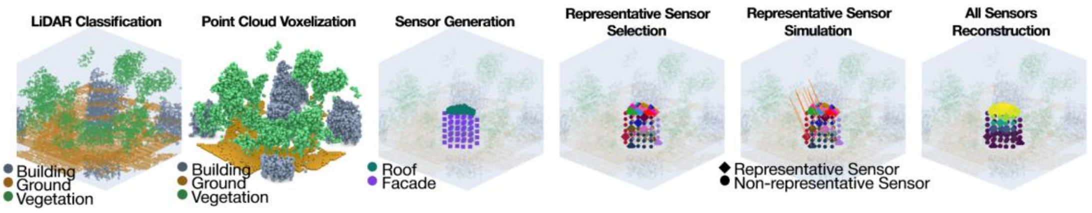

Figure 1*. ShadingZip workflow, selective computation of shading profiles*

Table 1. *Parameters for each feature family for representative sensor selection*

F	*Metadata features (MF)*	*Shading features (SF)*	*r*
SF	$\emptyset$	$\mathcal{E} \in \{\{9\},\{0,9\},\{0,4,9\}\}$; $m \in \{1,2,3\}$	$r \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.35, 0.5\}$
MF	Normal, tilt, roof/façade class, height above base, local surface (u,v)	$\mathcal{E} = \emptyset$; $m = 0$	
MF+SF		$\mathcal{E} \in \{\{9\},\{0,9\},\{0,4,9\}\}$; $m \in \{1,2,3\}$	

Table 2. *Representative sensor selection method*

Algorithm 1: Select representatives

Method SELECT-REPRESENTATIVES$(\mathcal{S}_b, \mathcal{E}, r, F)$
$K_b \leftarrow \min(|\mathcal{S}_b|, \max(1, \mathrm{round}(r|\mathcal{S}_b|)))$
Exploratory shading $Y_{\mathcal{E}} \leftarrow \mathrm{Raytrace}(\mathcal{S}_b, \mathcal{E})$ forming $\{\mathbf{f}_i\} \leftarrow \mathrm{Features}(\mathcal{S}_b, Y_{\mathcal{E}}, F)$
Clusters $\mathcal{C} \leftarrow \mathrm{KMeans}(\{\mathbf{f}_i\}, K_b)$ to initialize representatives $\mathcal{R}_b \leftarrow \emptyset$
for each centroid $\mathbf{c} \in \mathcal{C}$
  $i \leftarrow$ sensor nearest to $\mathbf{c}$ // get the real sensor closest to centroids
  while $i \in \mathcal{R}_b$ // check if the sensor is already selected (for another centroid)
    $i \leftarrow$ next-nearest sensor // if selected, use the next-nearest
  $\mathcal{R}_b \leftarrow \mathcal{R}_b \cup \{i\}$ // update representatives sensor set
for each $i \in \mathcal{S}_b$
  Representative sensor assigned $a(i) \leftarrow \arg\min_{j \in \mathcal{R}_b} \left\|\mathbf{f}_i - \mathbf{f}_j\right\|_2$
return $\mathcal{R}_b, a, Y_{\mathcal{E}}$

After representative selection, Algorithm 2 calculates the remaining shading periods only for sensors in $\mathcal{R}_b$. For an exploratory period, each sensor retains its own calculated profile. For non-exploratory periods, sensors copy the calculated profile of the assigned representative, yielding reconstructed profiles $\hat{Y}_b$. Buildings are streamed sequentially, posing no limit on scalability: once a building is processed, raw results are discarded and memory reclaimed.

Table 3. *Per-building computation and reconstruction workflow*

Algorithm 2: Compute and reconstruct profiles per building

for each building $b \in \mathcal{B}$ // streamed building access
  $\mathcal{R}_b, a, Y_\mathcal{E} \leftarrow$ `SELECT-REPRESENTATIVES`$(\mathcal{S}_b, \mathcal{E}, r, F)$ // run algorithm 1
  if $\mathcal{T} \setminus \mathcal{E} \neq \emptyset$ // check if temporal periods remain after excluding exploratory
    $Y_\mathcal{R} \leftarrow$ Raytrace$(\mathcal{R}_b, \mathcal{T} \setminus \mathcal{E})$ // trace only for representative sensors
  for each $i \in \mathcal{S}_b,\ t \in \mathcal{T}$ // for each sensor, for each period
    if $t \in \mathcal{E},\ \hat{y}_{i,t} \leftarrow Y_\mathcal{E}[i,t]$ // for an exploratory period, retain sensor's calculated result
    else $\hat{y}_{i,t} \leftarrow Y_\mathcal{R}[a(i),t]$ // for non-exploratory periods, use representative's result
  Save$(b, \mathcal{R}_b, a, \hat{Y}_b)$; Discard$(\hat{Y}_b)$; AutoGC() // saves results, frees up memory

We then explore the tradeoff between ray-tracing cost and reconstruction accuracy. Ray-tracing cost is measured relative to exhaustive calculation over all sensor periods. For exploratory periods, ray-tracing is performed for each sensor, whereas for non-exploratory periods, it is calculated for the representatives. Reconstruction accuracy is evaluated using a Weighted Mean Absolute Percentage Error (WMAPE):

$$\mathrm{WMAPE} = \frac{\sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}}\sum_{h\in\mathcal{H}} w_{i,t,h}\,|y_{i,t,h}-\hat{y}_{i,t,h}|}{\sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}}\sum_{h\in\mathcal{H}} w_{i,t,h}\,y_{i,t,h}},\quad w_{i,t,h} = A_i\,\pi_t\,(\sin\alpha_{t,h})_+(\mathbf{n}_i\cdot\mathbf{s}_{t,h})_+ \qquad (1)$$

Here, $y_{i,t,h}$ and $\hat{y}_{i,t,h}$ are ground-truth and reconstructed sunlit fractions for sensor i, period t, and hour h. A_i sensor i area; π_t is period t length; $\alpha_{t,h}$ is the solar altitude; $\mathbf{n}_i$ is the sensor normal; $\mathbf{s}_{t,h}$ is the direction vector to the sun, and $(\cdot)_+ = \max(0,\cdot)$. Unlit sensors were excluded as their percentage error is undefined. Profiles are aggregated for error calculation at surface- and building-levels, emulating how shading data enters BIPV and energy models. For baseline evaluations, sensors were randomly sampled within each building. All evaluations comprised 10 urban cases (***Table 4***) simulated at 0°, 15°, 30°, 45°, and 60° latitudes. All buildings were retained as obstruction context during shading calculation. A deterministic 10% sample, capped at 300 buildings per case, was selected for reconstruction and error evaluation. All computations were executed on a 24-core Intel Core Ultra 9 275HX processor and 32 GB RAM, with K-means restricted to 4 threads.

Table 4. *Test cases, contextual buildings (CB), selected buildings (SB) and sensors (SS) count*

Location	*Coordinates*	*Morphological characters*	*CB*	*SB*	*SS*
Financial District	40.71, -74.01	Urban, high-rise, street canyon	2,245	225	227,614
Midtown	40.75, -73.99	Urban, high-rise, grid	4,578	300	265,876
Red Hook	40.68, -74.00	Urban, industrial waterfront	6,719	300	102,846
Park Slope	40.67, -73.98	Urban, attached rowhouse	11,157	300	68,126
Riverdale	40.89, -73.91	Suburban, hilly, wooded	2,036	204	81,952
Co-op City	40.86, -73.83	Urban, towers, superblocks	3,698	300	61,399
Bayside	40.76, -73.77	Suburban, semi-detached	12,993	300	41,201
Jackson Heights	40.76, -73.88	Urban, perimeter blocks	17,761	300	60,092
Long Island City	40.75, -73.94	Urban, high-rise, industrial	3,149	300	226,362
Greenridge	40.55, -74.16	Suburban, residential	1,613	162	21,620
Total			65,949	2,691	1,157,088

RESULTS

The parameter sweep revealed a consistent tradeoff between ray-tracing cost and reconstruction error (Figure 2), with diminishing returns as the representative ratio increased. With the specified hardware, representative selection, reconstruction, and scoring for the selected SF configuration averaged 0.28 ms per sensor. Once at least 2 exploratory periods were accepted, SF and MF+SF generally outperformed the baseline. SF with $m = 3$ and $r = 0.20$ provided a preferred balance, requiring 32.6% of exhaustive sensor-period ray-tracing evaluations and producing a mean building-aggregated lit-sensor WMAPE of 2.6%. Across the 50 case–latitude combinations, WMAPE ranged from 0.7% to 5.5%, with 49 combinations below 5% (Figure 4). Figure 3 shows that facade-surface error is broadly distributed across orientations at 0° latitude but becomes more concentrated on north-facing facades at higher latitudes. Although the local facade error exceeds the building-aggregated error, the selected configuration retains low-error surfaces across all orientations.

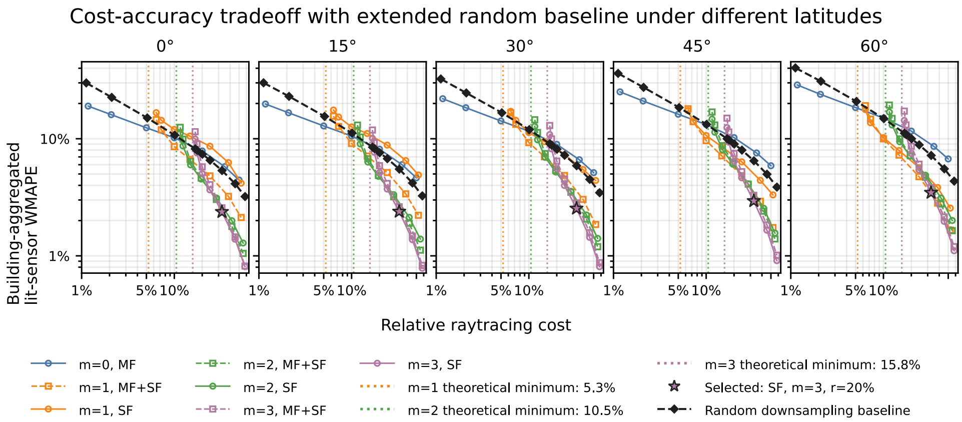

***Figure 2**. Cost-accuracy tradeoff for each feature family under different latitudes*

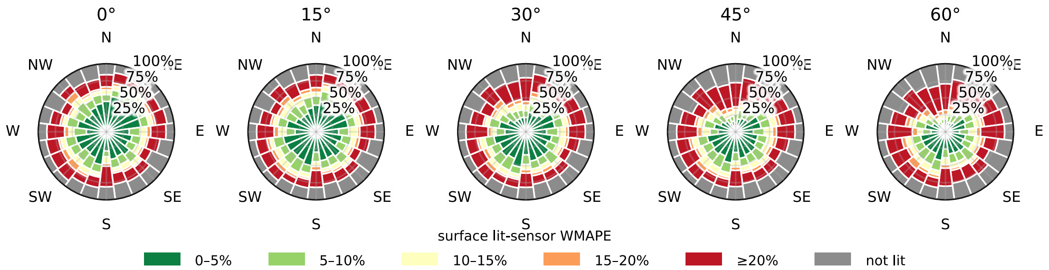

***Figure 3**. Facade lit-sensor WMAPE for SF (m=3, r=0.2) by latitude and orientation*

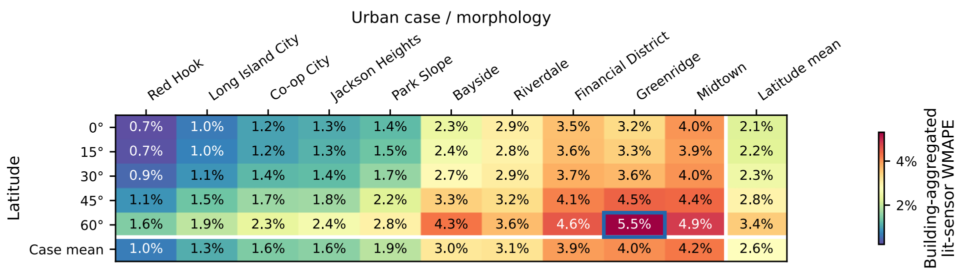

***Figure 4**. Building-aggregated lit-sensor WMAPE for SF (m=3, r=0.2) across ten urban cases and five latitudes.*

DISCUSSION

Exploratory shading observations capture obstruction-dependent temporal structure that geometry alone cannot reliably encode. Consequently, SF and MF+SF outperform MF once one exploratory period is affordable, suggesting MF may introduce variation unrelated to shading-profiles. SF with $m = 3$ and $r = 0.20$ provides the clearest balance between accuracy and ray-tracing workload. The matched-cost random downsampling baseline confirms that the improvement is not explained by reduced evaluated sensor count alone. Exploratory shading directs computation toward sensors whose temporal profiles better represent the remaining sensors, yielding approximately 2.2× lower WMAPE. Moreover, random downsampling estimates aggregate building profiles, whereas ShadingZip reconstructs outputs for every sensor and retains building and surface identities, and is particularly useful for location-based queries adopted in window or BIPV models. Orientation-dependent errors qualify the strong building-level performance. North-facing facades exhibit higher surface errors at higher latitudes as their smaller signal produces a sensitive WMAPE denominator, yielding limited contribution to building-level error. The nearly uniform orientation pattern at 0° suggests little directional bias in the evaluated sample. Meanwhile, ShadingZip addresses a different aspect of computational cost from pixel-counting, ray-tracing, or voxel acceleration methods. Those methods reduce the cost of each evaluation, whereas ShadingZip reduces the number of evaluations requested. They are therefore complementary and may provide compound savings when combined.

CONCLUSION AND IMPLICATIONS

ShadingZip reduces shading computation by tracing complete profiles only for representative sensors selected within each building. Across ten morphologies and five latitudes, SF with $m = 3$ and $r = 0.20$ achieved a mean building-aggregated lit-sensor WMAPE of 2.6% saving 67.4% ray-tracing effort. Future work should determine whether the same selective representation can jointly reconstruct sky-patch visibility or sky-view factors. Meanwhile, ShadingZip still depends on exploratory ray-tracing; future ray-tracing-free methods could learn conditional distributions of shading profiles from geometry and context direcly.

REFERENCES

Hong, Tianzhen, Yixing Chen, Xuan Luo, Na Luo, and Sang Hoon Lee. 2020. "Ten Questions on Urban Building Energy Modeling." Building and Environment 168 (January): 106508.

Hoover, Joel, and Timur Dogan. 2017. "Fast and Robust External Solar Shading Calculations Using the Pixel Counting Algorithm with Transparency." Building Simulation, vol. 15: 1083–91.

Kent, Michael G., Stefano Schiavon, and John Alstan Jakubiec. 2020. "A Dimensionality Reduction Method to Select the Most Representative Daylight Illuminance Distributions." Journal of Building Performance Simulation 13 (1): 122–35.

Luo, Xuan, Yu-Hang Tang, and Tianzhen Hong. 2020. "Efficient Computation of Surface Sunlit Fractions in Urban-Scale Building Modeling Using Ray-Tracing Techniques." SimBuild Conference, vol. 9: 238–43.

Su, Amber Jiayu, and Timur Dogan. 2026. "Voxel Shader – An Efficient Shadow Calculation Approach for Urban Building Energy Modeling." Energy and Buildings 357 (April): 117135.