

Conflated Inverse Modeling to Generate Diverse and Temperature-Change Inducing Urban Vegetation Patterns

Supplementary Material

S1. Additional Dataset-Level Analysis

We provide an additional dataset-level analysis to illustrate the one-to-many nature of the inverse problem. Specifically, we measure tile-level NDVI variability after grouping tiles into similar building-height and mean-LST bins.

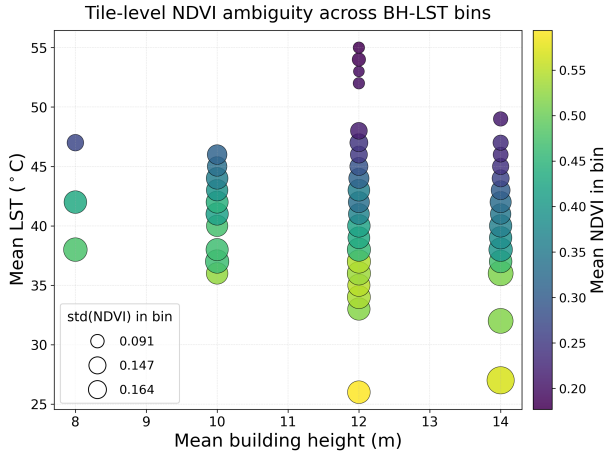


Figure 1. **NDVI Diversity.** We show tile-level NDVI variability observed in our dataset after binning tiles into similar building height and mean LST bins. Marker size corresponds to the standard deviation of NDVI within each bin (minimum count = 30). Standard deviations range from 0.09 to 0.16, indicating substantial within-bin variability.

S2. Additional Qualitative Results Across Cities

We provide additional qualitative results across multiple cities in our dataset, shown in Fig. 2. For each city, we show NDVI generation within the ROI under $\Delta \in \{-2, 0, +2\}^\circ\text{C}$. The reported values correspond to the ROI mean predicted temperature change Δ_{pred} obtained from the frozen forward model.

The 20 cities included in our dataset are: Billings, MT; Bloomington, IN; Buffalo, NY; Chicago, IL; Denver, CO; Eugene, OR; Green Bay, WI; Indianapolis, IN; Jacksonville, FL; Los Angeles, CA; Norman, OK; Overland Park, KS; Phoenix, AZ; Pittsburgh, PA; Reno, NV; Salt Lake City, UT; San Jose, CA; Shreveport, LA; Sioux Falls, SD; and Syracuse, NY.

S3. Implementation Details

Forward model implementation details. The forward predictor g_ϕ is implemented as a U-Net using

`segmentation_models_pytorch` with a ResNet-50 encoder and ImageNet-pretrained encoder weights. We train the model for 15 epochs with batch size 32 using Adam, a learning rate of 3×10^{-4} , weight decay 10^{-4} , and an MSE loss. During training, we apply an urban-weighted per-pixel loss with weight 5.0 on urban pixels, where the urban mask is used only for loss weighting. Urban pixels are defined as locations with positive building-height values. The best checkpoint is selected based on validation MAE on urban pixels, and the trained forward model is then frozen during inverse model training.

End-to-end inverse U-Net baseline. The end-to-end inverse U-Net baseline uses the same `segmentation_models_pytorch` U-Net architecture as the forward predictor, with a ResNet-50 encoder and ImageNet-pretrained encoder weights. The model is trained for 20 epochs with batch size 32 using Adam, a learning rate of 3×10^{-4} , no weight decay, and an L1 loss. The best checkpoint is selected based on NDVI MAE on the validation set, with the best model obtained at epoch 19.

Inverse diffusion model implementation details. The full architecture, optimization, training, and sampling hyperparameters are summarized in Table 1.

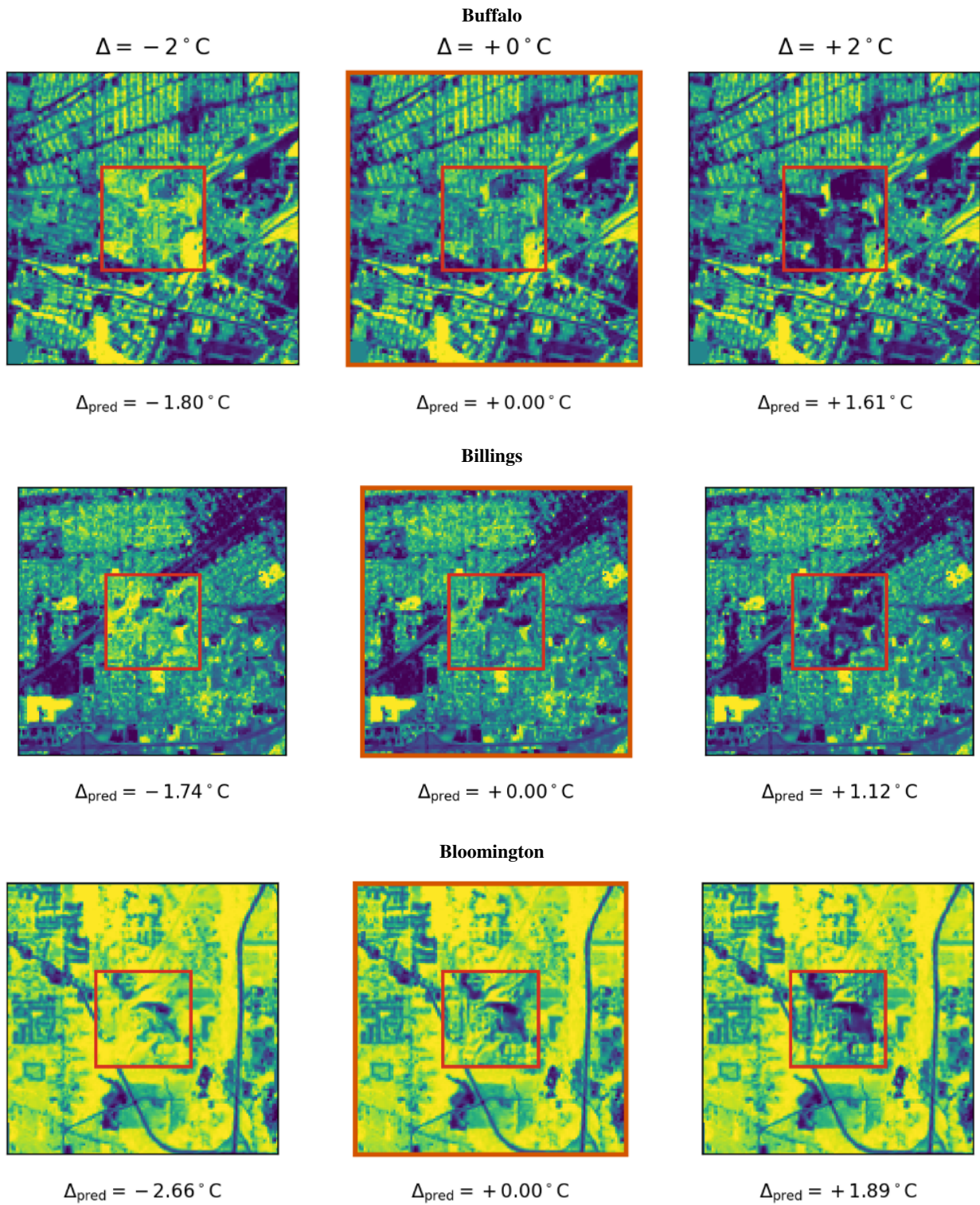
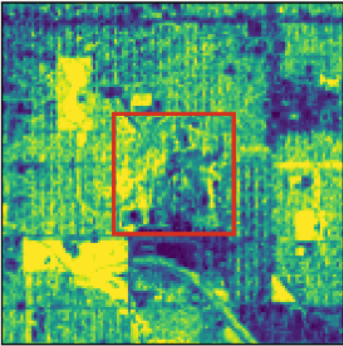
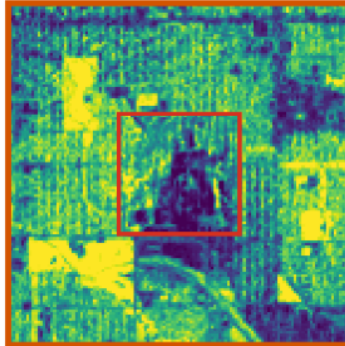


Figure 2. Additional city-level qualitative results. Each row shows NDVI generation under $\Delta \in \{-2, 0, +2\}^{\circ}\text{C}$ with the ROI mean predicted change Δ_{pred} .

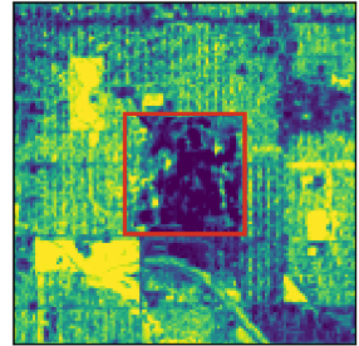
Denver



$\Delta_{\text{pred}} = -1.57^\circ\text{C}$

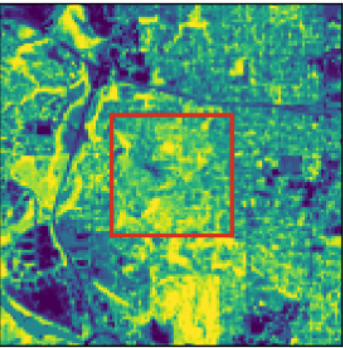


$\Delta_{\text{pred}} = +0.00^\circ\text{C}$

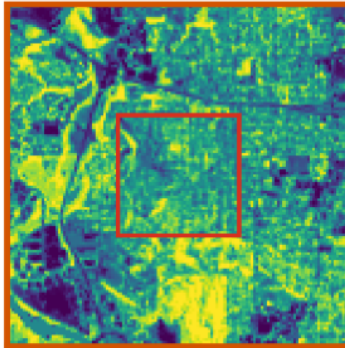


$\Delta_{\text{pred}} = +1.30^\circ\text{C}$

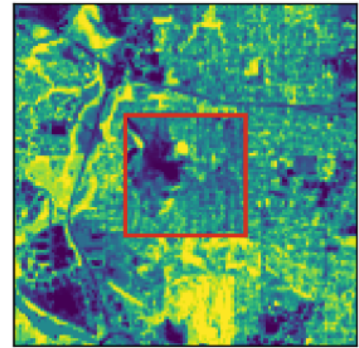
Eugene



$\Delta_{\text{pred}} = -1.88^\circ\text{C}$

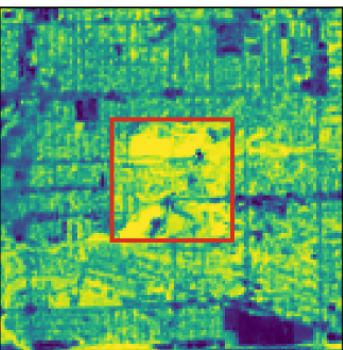


$\Delta_{\text{pred}} = +0.00^\circ\text{C}$

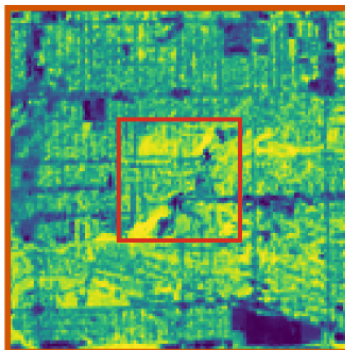


$\Delta_{\text{pred}} = +1.56^\circ\text{C}$

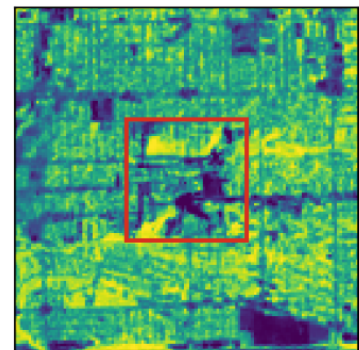
Indianapolis



$\Delta_{\text{pred}} = -2.13^\circ\text{C}$



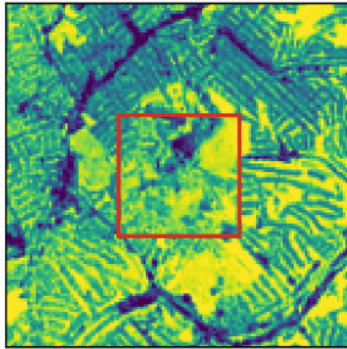
$\Delta_{\text{pred}} = +0.00^\circ\text{C}$



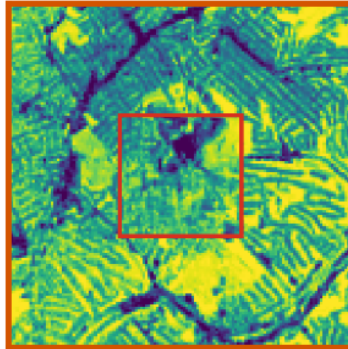
$\Delta_{\text{pred}} = +2.30^\circ\text{C}$

Figure 2 (continued).

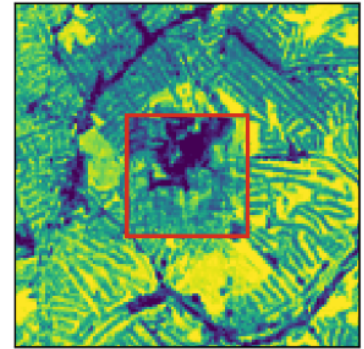
Pittsburgh



$\Delta_{\text{pred}} = -1.98^\circ\text{C}$

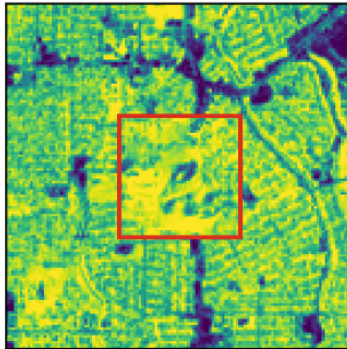


$\Delta_{\text{pred}} = +0.00^\circ\text{C}$

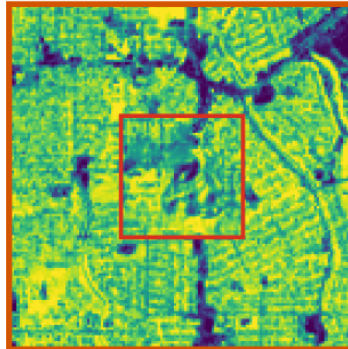


$\Delta_{\text{pred}} = +1.94^\circ\text{C}$

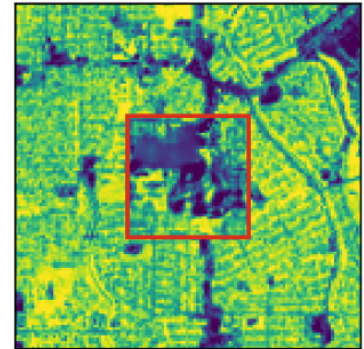
Shreveport



$\Delta_{\text{pred}} = -2.59^\circ\text{C}$



$\Delta_{\text{pred}} = +0.00^\circ\text{C}$

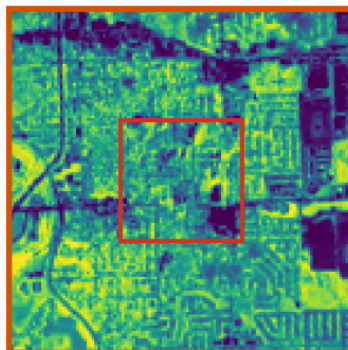


$\Delta_{\text{pred}} = +1.93^\circ\text{C}$

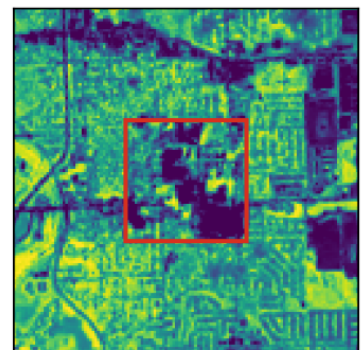
Sioux Falls



$\Delta_{\text{pred}} = -2.02^\circ\text{C}$



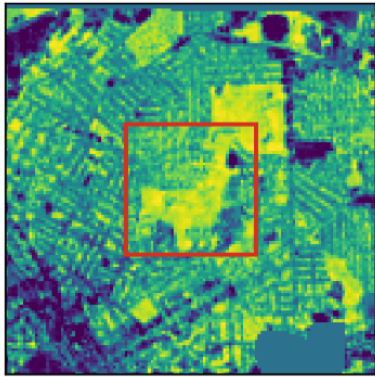
$\Delta_{\text{pred}} = +0.00^\circ\text{C}$



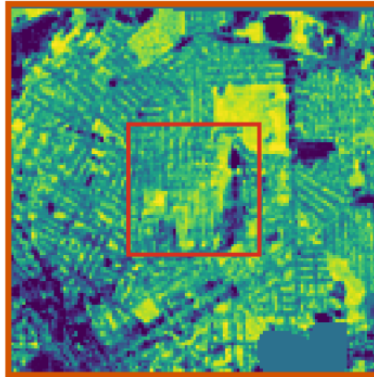
$\Delta_{\text{pred}} = +2.39^\circ\text{C}$

Figure 2 (continued).

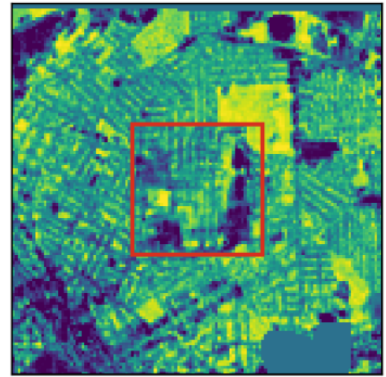
Syracuse



$\Delta_{\text{pred}} = -1.49^\circ\text{C}$



$\Delta_{\text{pred}} = +0.00^\circ\text{C}$



$\Delta_{\text{pred}} = +1.53^\circ\text{C}$

Figure 2 (continued).

Table 1. Inverse diffusion model hyperparameters.

Parameter	Value
Architecture	
Backbone	NCSN++
Image size	128×128
Base channels (nf)	128
Channel multipliers	(1, 2, 2, 2)
Residual blocks / resolution	4
Attention resolutions	(16)
Normalization	GroupNorm
Nonlinearity	swish
Residual block type	BigGAN
Optimization	
Batch size	16
Optimizer	Adam
Learning rate	2×10^{-4}
Adam β_1	0.9
Gradient clipping	1.0
LR warmup steps	5000
EDM	
σ_{data}	0.5
P_{mean}	-1.2
P_{std}	1.2
Physics loss	
Physics loss	L1
Physics warmup steps	5000
Physics ramp steps	10000
Sampling	
Sampling method	EDM
Sampler steps	40
ρ	7.0
σ_{min}	0.002
σ_{max}	80.0