# VIBRANTSR: SUB-METER CANOPY HEIGHT MODELS FROM SENTINEL-2 USING GENERATIVE FLOW MATCHING

**Kiarie Ndegwa,*** **Andreas Gros,†** **Tony Chang, David Diaz, Vincent A. Landau,**
Nathan E. Rutenbeck, Luke J. Zachmann, Guy Bayes,† Scott Conway

*Vibrant Planet Public Benefit Corporation, Truckee, CA, USA*

## ABSTRACT

We present VibrantSR, a generative super-resolution framework that estimates 0.5 m canopy height models (CHMs) from 10 m Sentinel-2 imagery using latent-space flow matching. Evaluated across 22 EPA Level 3 eco-regions in the western United States, VibrantSR achieves a Mean Absolute Error of 4.39 m for canopy heights $\geq$2 m, outperforming satellite-based benchmarks Meta (4.83 m), LANDFIRE (5.96 m), and ETH (7.05 m) with lower edge error. An aerial-imagery baseline (2.71 m MAE) retains an accuracy advantage, reflecting the fundamental resolution gap between 10 m and 0.5 m inputs. VibrantSR offers a scalable alternative for regional forest monitoring that improves on existing satellite-derived CHM products while avoiding reliance on costly aerial acquisitions.

## 1 INTRODUCTION

High-resolution canopy height models (CHMs) support wildfire planning, carbon accounting, and habitat assessment (North et al., 2021; Cruz et al., 2005; Coops et al., 2021). Airborne lidar provides accurate CHMs but is costly and temporally sparse (Bolton et al., 2020). Aerial imagery-based approaches such as those using NAIP can achieve sub-meter accuracy, but are constrained by irregular acquisition schedules, infrequent revisit cadence (typically 2–3 years), and radiometric inconsistencies across flight paths—limiting suitability for operational monitoring at continental scales. Sentinel-2 offers global, repeatable observations at seasonal cadence but at 10–60 m resolution, insufficient to resolve sub-meter canopy structure (Drusch et al., 2012).

We introduce **VibrantSR**, a generative framework that produces 0.5 m CHMs from 10 m Sentinel-2 imagery using flow matching in a compressed latent space. Unlike regression-based methods that smooth predictions toward the mean, our generative formulation preserves realistic canopy variability critical for applications such as crown delineation and gap mapping. We evaluate on the western US across 22 EPA Level 3 eco-regions (U.S. Environmental Protection Agency, 2013), benchmarking against satellite-derived CHM products (Tolan et al., 2024; LANDFIRE, 2022; Lang et al., 2023) and an aerial-imagery baseline (a vision transformer trained on 0.5 m NAIP imagery with lidar supervision; Chang et al. 2025).

## 2 RELATED WORK

Deep learning CHM estimation has evolved from CNNs to vision transformers (Tolan et al., 2024; Lang et al., 2023; Wagner et al., 2024). Our method builds on flow matching generative models (Lipman et al., 2023; Liu et al., 2023; Esser et al., 2024) and remote sensing super-resolution (Sirko et al., 2023; Aybar et al., 2024; Lanaras et al., 2018). Generative downscaling of geophysical fields has been explored in climate applications (Harris et al., 2022; Price & Rasp, 2022). We are, to our knowledge, the first to apply latent flow matching to super-resolve fine-scale canopy structure from multispectral satellite imagery.

*Corresponding author: kiarie@vibrantplanet.net
†Work carried out while at Vibrant Planet.

Table 1: Baseline input data and output resolution.

| Model | Resolution | Input Data | Coverage |
|---|---|---|---|
| Aerial baseline | 0.5 m | NAIP aerial | Western US |
| Meta | 1 m | Maxar Vivid2 | Global |
| LANDFIRE | 30 m | Landsat | United States |
| ETH | 10 m | Sentinel-2 | Global |

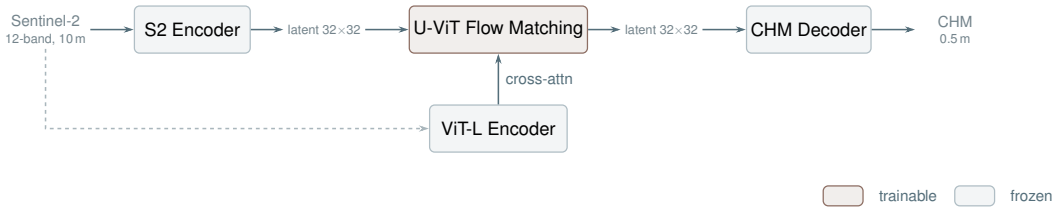


Figure 1: VibrantSR architecture: frozen autoencoders compress Sentinel-2 inputs and decode CHM outputs; a trainable U-ViT flow matching network, conditioned on self-supervised ViT-L features via cross-attention, maps between latent spaces.

## 3 Data

We use 12-band Sentinel-2 Level-2A surface reflectance (Drusch et al., 2012; Element 84, 2023), aggregated via temporal median over summer months (June–August 2024) after cloud masking. Each sample pairs a $48\times48$ Sentinel-2 tile ($480\,\mathrm{m}\times480\,\mathrm{m}$ at 10 m) with a corresponding $960\times960$ CHM tile at 0.5 m derived from USGS 3DEP airborne lidar (U.S. Geological Survey, 2024). Metrics are reported for canopy heights $\geq$2 m (Van Leeuwen & Nieuwenhuis, 2010). The dataset contains 168,834 training and 66,154 validation tiles over 4,135 km$^2$ spanning 22 EPA Level 3 ecoregions in the western US (U.S. Environmental Protection Agency, 2013), using spatially disjoint checkerboard splits following Chang et al. (2025).

## 4 Methods

VibrantSR (Fig. 1) operates in compressed latent space with three components: (1) a frozen Sentinel-2 autoencoder; (2) a trainable U-ViT flow matching network (16 transformer layers, 16 attention heads, $32\times32$ latent grids), conditioned via cross-attention on a frozen self-supervised ViT-L encoder; and (3) a frozen CHM decoder. At inference, $\frac{dz}{dt}=v_\theta(z,t)$ is integrated from $t$=0 to $t$=1 using `dopri5` with 100 steps (Lipman et al., 2024; Chen et al., 2018). The network was trained for 128 epochs using AdamW (lr $=10^{-4}$, cosine decay, weight decay $=0.03$, batch size 12) on 8$\times$A100 GPUs.

We compare against four baselines and report Mean Absolute Error (MAE), Mean Error (ME, systematic bias), and Edge Error (EE). EE measures structural fidelity as the mean absolute difference between Sobel-filtered predictions and references: $\mathrm{EE}=\frac{1}{n}\sum_i|S(\hat{y}_i)-S(y_i)|$, where $S(\cdot)$ is the Sobel edge operator. Lower EE indicates sharper, better-localized canopy boundaries—critical for downstream tasks like crown delineation and gap mapping (Wagner et al., 2024). Baseline models span different input modalities and resolutions (Table 1).

## 5 Results

### 5.1 Quantitative Comparison

VibrantSR achieves MAE of 4.39 m, outperforming all satellite-based models: 9% lower than Meta (4.83 m), 26% lower than LANDFIRE (5.96 m), and 38% lower than ETH (7.05 m; Table 2). Edge error is also substantially lower, indicating superior structural fidelity. Qualitative comparisons (Fig. 2) show VibrantSR recovers fine-scale canopy structure from coarse Sentinel-2 inputs that is typically smoothed by regression-based approaches.

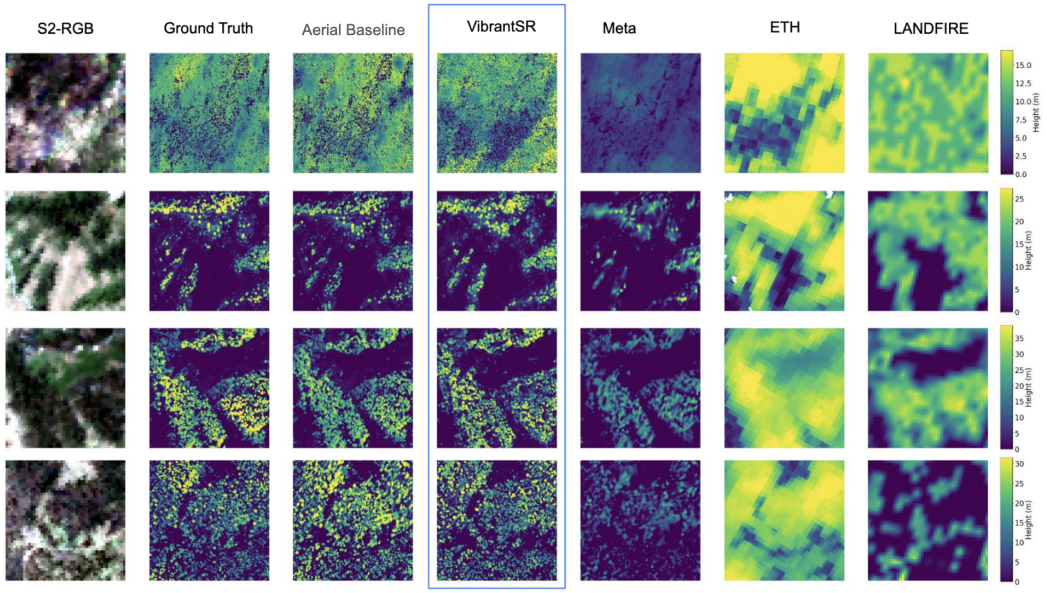


Figure 2: Qualitative comparison across multiple regions (one per row). VibrantSR generates plausible sub-meter canopy structure from 10 m Sentinel-2 inputs, more closely matching lidar reference than satellite-based competitors.

Table 2: Performance at 0.5 m for heights $\geq$2 m. Bold = best satellite-only model; the aerial baseline uses 0.5 m NAIP inputs.

| Model | MAE | ME | EE |
|---|---|---|---|
| **VibrantSR** | **4.39** | **-2.35** | **0.16** |
| Aerial baseline | 2.71 | -1.11 | 0.08 |
| Meta | 4.83 | -4.03 | 0.30 |
| LANDFIRE | 5.96 | 0.92 | 0.63 |
| ETH | 7.05 | 5.65 | 0.64 |

### 5.2 Ablation: Flow Matching vs. Regression

Due to compute constraints, we ran ablations on a smaller subset: both variants were trained on the same 10k tiles and evaluated on the same held-out 2k tiles. Table 3 compares flow matching against MSE regression with identical architecture. The flow model achieves lower edge error (0.29 vs. 0.37) but higher MAE (4.47 vs. 2.40 m)—a characteristic tradeoff where MSE regression produces blurry outputs that smear canopy boundaries (Wang et al., 2020), while flow matching preserves structural detail at the cost of pixel accuracy. For crown delineation and gap detection, accurate edge localization is essential (Wagner et al., 2024).

Table 3: Ablation: flow matching vs. MSE regression on a held-out test set (n=2,000 tiles, non-overlapping with the 10k training set). MAE measures mean absolute error in meters (lower is better); EE measures edge error (lower is better). Regression achieves lower MAE but worse edge fidelity.

| Variant | MAE ↓ | EE ↓ |
|---|---|---|
| VibrantSR (flow matching) | 4.47 | **0.29** |
| VibrantSR (MSE regression) | **2.40** | 0.37 |

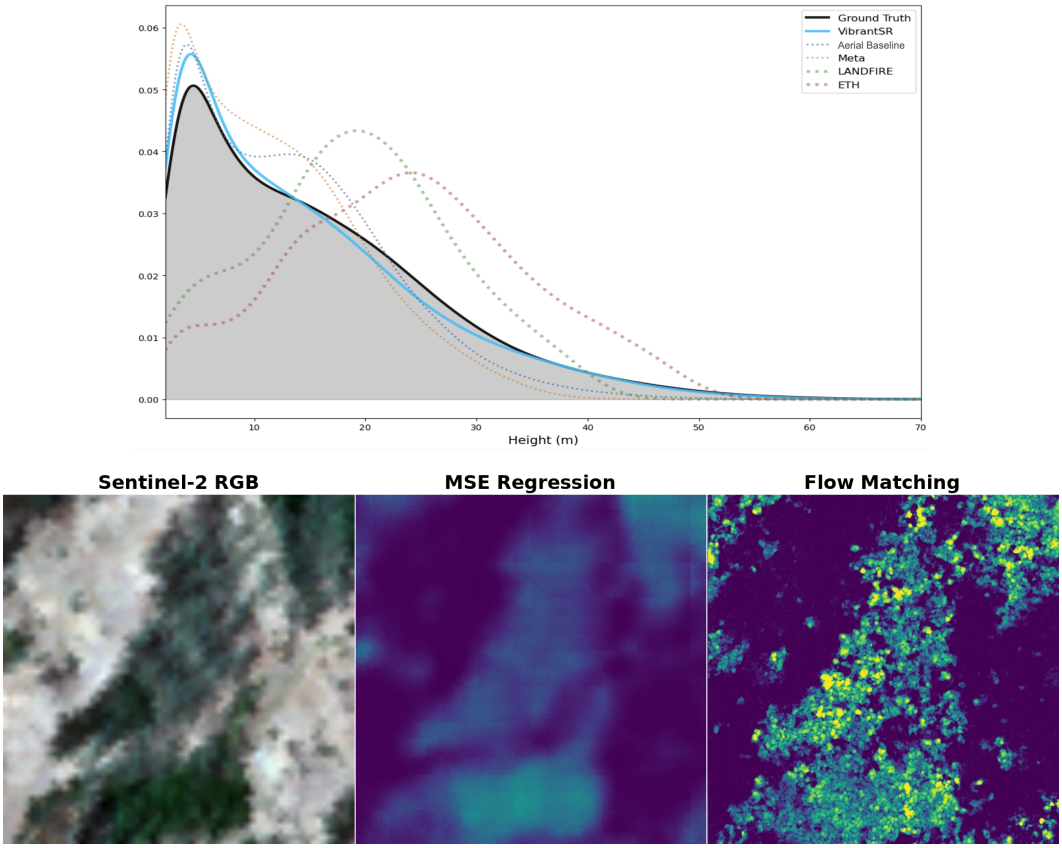


Figure 3: Top: kernel density estimates of canopy height distributions. VibrantSR (blue) matches the lidar-derived reference distribution more closely than satellite baselines. Bottom: qualitative comparison of flow matching vs. MSE regression. Left: Sentinel-2 RGB. Center: regression produces L2 blur. Right: flow matching recovers sharp canopy boundaries.

## 6 Discussion and Conclusion

VibrantSR produces 0.5 m CHMs from 10 m Sentinel-2 using latent-space flow matching, reducing MAE by 9–38% relative to satellite-derived baselines with lower edge error. The generative formulation preserves canopy-height distributions that regression-based methods tend to smooth (Fig. 3, top), a property critical for downstream applications requiring realistic structural variability.

The aerial baseline outperforms VibrantSR (2.71 vs. 4.39 m MAE), reflecting the fundamental information gap between 10 m and 0.5 m inputs. VibrantSR occupies a distinct niche: for project-level planning requiring highest accuracy, aerial-based methods or direct lidar remain preferable. However, for regional screening and prioritization where broad coverage matters more than sub-meter precision, VibrantSR offers a practical solution. The 20× upsampling is mediated by the frozen CHM decoder, so sub-meter detail should be interpreted as plausible structure rather than measured geometry. Current evaluation is limited to the western US; generalization to other regions and biomes with different canopy structures remains an important open question. Due to compute constraints, we report only deterministic outputs; future work will leverage stochastic sampling to generate per-pixel confidence intervals and explore auxiliary modalities (e.g., SAR).

## References

Cesar Aybar, David Montero, Simon Donike, Freddie Kalaitzis, and Luis Gómez-Chova. A comprehensive benchmark for optical remote sensing image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024. doi: 10.1109/LGRS.2024.3401394.

Douglas K. Bolton, Nicholas C. Coops, Michael A. Wulder, Joanne C. White, Txomin Hermosilla, and Adam Collingwood. Optimizing landsat time series length for regional mapping of lidar-derived forest structure. *Remote Sensing of Environment*, 239:111645, 2020.

Tony Chang, Kiarie Ndegwa, Andreas Gros, Vincent A. Landau, Luke J. Zachmann, B. State, M.A. Gritts, C.W. Miller, Nathan E. Rutenbeck, Scott Conway, and Guy Bayes. Vibrantvs: A high-resolution vision transformer for forest canopy height estimation. *Remote Sensing*, 17(6):1017, 2025. doi: 10.3390/rs17061017.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, 2018. URL https://arxiv.org/abs/1806.07366.

Nicholas C. Coops, Piotr Tompalski, Tristan R.H. Goodbody, Martin Queinnec, Joan E. Luther, Douglas K. Bolton, Joanne C. White, Michael A. Wulder, Oliver R. van Lier, and Txomin Hermosilla. Modelling lidar-derived estimates of forest attributes over space and time: A review of approaches and future trends. *Remote Sensing of Environment*, 260:112477, 2021.

Miguel G. Cruz, Martin E. Alexander, and Ronald H. Wakimoto. Development and testing of models for predicting crown fire rate of spread in conifer forest stands. *Canadian Journal of Forest Research*, 35(7):1626–1639, 2005.

Matthias Drusch, Umberto Del Bello, Sebastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote Sensing of Environment*, 120:25–36, 2012.

Element 84. Earth search stac api: Sentinel-2 collection 1 level-2a (sentinel-2-c1-l2a). https://github.com/Element84/earth-search, 2023. README documentation describing the Earth Search STAC API and the `sentinel-2-c1-l2a` collection. Accessed 2026-01-08.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning (ICML)*, 2024.

Lucy Harris, Andrew T. T. McRae, Matthew Chantry, Peter D. Dueben, and Tim N. Palmer. A generative deep learning approach to stochastic downscaling of precipitation forecasts. *Journal of Advances in Modeling Earth Systems*, 14(10):e2022MS003120, 2022. doi: 10.1029/2022MS003120.

Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:305–319, 2018.

LANDFIRE. Landfire fuels - forest canopy height. Technical report, U.S. Geological Survey, 2022.

N. Lang, W. Jetz, K. Schindler, and J.D. Wegner. A high-resolution canopy height model of the earth. *Nature Ecology & Evolution*, 7:1778–1789, 2023.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code, 2024. URL https://arxiv.org/abs/2412.06264.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023.

Malcolm P. North, Robert A. York, Brandon M. Collins, Matthew D. Hurteau, Gavin M. Jones, Eric E. Knapp, Leda Kobziar, H. Millard, et al. Pyrosilviculture needed for landscape resilience of dry western united states forests. *Journal of Forestry*, 119(5):520–544, 2021.

Ilan Price and Stephan Rasp. Increasing the accuracy and resolution of precipitation forecasts using deep generative models. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 151 of *Proceedings of Machine Learning Research*, pp. 10555–10571. PMLR, 2022.

Wojciech Sirko, Emmanuel Asiedu Brempong, Juliana TC Marcos, Abigail Annkah, Abel Korme, Mohammed Alewi Hassen, Krishna Sapkota, Tomer Shekel, Abdoulaye Diack, Sella Nevo, Jason Hickey, and John Quinn. High-resolution building and road detection from sentinel-2. *arXiv preprint arXiv:2310.11622*, 2023.

J. Tolan, H.-I. Yang, B. Nosarzewski, et al. Very high resolution canopy height maps from rgb imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300:113888, 2024.

U.S. Environmental Protection Agency. Level iii ecoregions of the continental united states. Technical report, U.S. Environmental Protection Agency, 2013.

U.S. Geological Survey. 3d elevation program (3dep). Technical report, U.S. Geological Survey, 2024.

Martin Van Leeuwen and Maarten Nieuwenhuis. Retrieval of forest structural parameters using lidar remote sensing. *European Journal of Forest Research*, 129(4):749–770, 2010.

Fabien H. Wagner, Sophia Roberts, Alison L. Ritz, Griffin Carter, Ricardo Dalagnol, Samuel Favrichon, Mayumi C.M. Hirye, Martin Brandt, Philippe Ciais, and Sassan Saatchi. Sub-meter tree height mapping of california using aerial images and lidar-informed u-net model. *Remote Sensing of Environment*, 305:114099, 2024.

Zhihao Wang, Jian Chen, and Steven C. H. Hoi. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3365–3387, 2020.