Physics-Constrained Symbolic Regression from Imagery

Zhenyu Yu [1 2] **Mohd Yamani Idna Idris** [2] **Pei Wang** [1 *]

Abstract

We propose *SymbolicVision*, a physics-constrained symbolic regression framework that derives interpretable mathematical expressions directly from multi-spectral remote sensing imagery. Unlike black-box deep models, *SymbolicVision* combines a Vision-based image encoder with a Transformer-based symbolic decoder to enable cross-modal learning between visual features and symbolic formulas. A hybrid loss design ensures both numerical accuracy and physical plausibility. Evaluated on symbolic benchmarks (SRBench) and real satellite datasets (Open-Canopy), *SymbolicVision* achieves high predictive accuracy ($R^2 > 0.99$), and robust performance on geospatial tasks. This work highlights the potential of interpretable, physics-aware models for scientific remote sensing.

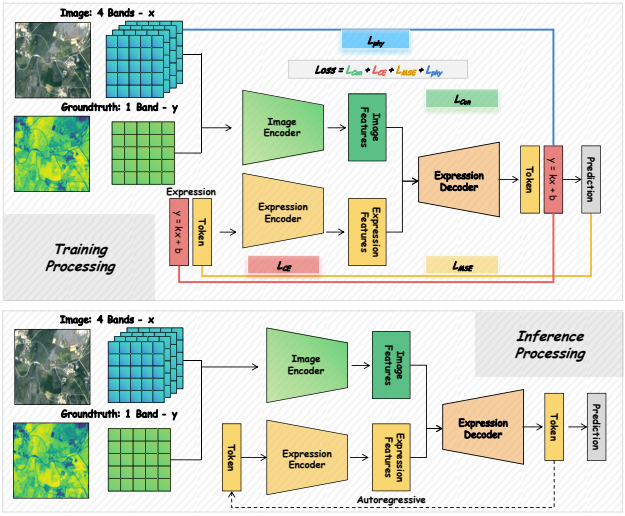

Figure 1. Architecture of *SymbolicVision*. The framework integrates a vision encoder, symbolic decoder, and cross-modal fusion under physics-based supervision.

1. Introduction and Motivation

Remote sensing imagery captures essential physical and environmental information across multiple spectral bands, such as vegetation health, soil moisture, and surface temperature (Zhu et al., 2017; Yu et al., 2024; 2025d; Yu, 2025; Makke & Chawla, 2024a; Grayeli et al., 2024). However, extracting physically interpretable expressions from such data remains a major challenge (AbdusSalam et al., 2025; Makke & Chawla, 2024b; Merler et al., 2024). Traditional empirical methods, like the Normalized Difference Vegetation Index (NDVI) and the Soil-Adjusted Vegetation Index (SAVI), rely on predefined formulas that require extensive domain knowledge and verification (Huete et al., 2018; Davis, 2024; Yu et al., 2025a;e). Machine learning approaches (e.g., Random Forest, UNet, Transformers) improve accuracy but offer limited interpretability (Chen & Guestrin, 2016; Zhu et al., 2017; Zhou et al., 2024; Wang et al., 2024; Yu et al., 2025b). **Symbolic regression (SR)** offers an interpretable alternative, discovering human-readable formulas from data (Koza, 1994; Udrescu & Tegmark, 2020; Tashtoush et al., 2024; Zhang et al., 2024a; Yu et al., 2025c; Tian et al., 2025; Zhang et al., 2024b). Yet, most SR methods are built for tabular data and fail on high-dimensional spatial-spectral imagery. Transformer-based SR (e.g., SymbolicGPT (Kamienny et al., 2021), NeSymReS (Biggio et al., 2021), TPSR (Landajuela et al., 2023)) scales better but lacks support for image data and physics constraints. **Challenges** remain in extracting symbolic knowledge directly from remote sensing imagery due to complex spatial dependencies and spectral variance (Jiang et al., 2024; Hochmair et al., 2025; Glazer et al., 2024; Yang et al., 2024; Polydoros et al., 2025). We propose *SymbolicVision*, a vision-symbolic framework combining Swin Transformer encoders and symbolic decoders, with physics-informed constraints. This enables accurate, interpretable expressions from imagery, bridging remote sensing and scientific discovery.

*Corresponding author [1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China [2]Department of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, Malaysia. Correspondence to: Pei Wang <peiwang@kust.edu.cn>, Zhenyu Yu <yuzhenyuyxl@foxmail.com>.

The second AI for MATH Workshop at Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

Table 1. Performance comparison of different models. The **bold** is the best, while the underlined is the second. N denotes the nodes.

Dataset	*SymbolicVision*		MMSR		TPSR		End2End		NeSymReS		SymbolicGPT	
	R^2↑	N↓	R^2↑	N↓	R^2↑	N↓	R^2↑	N↓	R^2↑	N↓	R^2↑	N↓
Nguyen	0.9996±0.004	**14.5**	**0.9999±0.001**	**14.5**	0.9948±0.002	16	0.8814±0.004	16.3	0.8568±0.003	18.2	0.6713±0.005	21.6
Keijzer	0.9980±0.004	**16.3**	**0.9983±0.003**	**16.3**	0.9828±0.003	20.6	0.8134±0.005	18.4	0.7992±0.003	21.3	0.6031±0.004	24.5
Korns	0.9979±0.004	**19.2**	**0.9982±0.003**	**19.2**	0.9325±0.004	22.9	0.8715±0.004	23.4	0.8011±0.005	24.1	0.6613±0.005	29.2
Constant	0.9983±0.004	**24.5**	**0.9986±0.002**	**24.5**	0.9319±0.002	35.3	0.8015±0.003	28.3	0.8344±0.003	32.9	0.7024±0.004	38.5
Livermore	0.9815±0.005	**29.4**	**0.9844±0.003**	**29.4**	0.882±0.004	38.2	0.7015±0.004	32.2	0.6836±0.005	36.2	0.5631±0.0005	41.2
Vladislavleva	0.9859±0.004	**21.7**	**0.9862±0.003**	**21.7**	0.9028±0.005	24.6	0.7422±0.005	22.2	0.6892±0.004	27.3	0.5413±0.004	36.6
R	0.9918±0.005	16.4	**0.9924±0.004**	16.4	0.9422±0.003	**16.2**	0.8512±0.004	19.5	0.7703±0.005	19.9	0.7042±0.005	25.2
Jin	0.9937±0.004	**28.3**	**0.9943±0.003**	**28.3**	0.9826±0.004	29.5	0.8611±0.004	29.8	0.8327±0.003	32.2	0.7724±0.006	36.9
Neat	0.9969±0.005	17.3	**0.9972±0.004**	17.3	0.9319±0.002	**16.4**	0.8044±0.004	19.7	0.7596±0.005	20.6	0.6377±0.005	26.4
Others	0.9985±0.004	**20.6**	**0.9988±0.002**	**20.6**	0.9667±0.002	22.5	0.8415±0.003	22.3	0.8026±0.003	23.5	0.7031±0.004	31.8
Feynman	0.9908±0.005	**20.8**	**0.9913±0.002**	**20.8**	0.8928±0.004	21.3	0.7353±0.004	22	0.7025±0.005	22.4	0.5377±0.005	26.8
Strogatz	0.9789±0.005	**21.6**	**0.9819±0.003**	**21.6**	0.8249±0.002	24.4	0.6626±0.003	25.4	0.6022±0.003	28.1	0.5229±0.004	32.6
Black-box	0.9934±0.005	**26.7**	**0.9937±0.004**	**26.7**	0.8753±0.004	29.3	0.6925±0.004	31.2	0.6525±0.005	33.9	0.5833±0.005	37.4
Average	0.993	**21.3**	**0.9934**	**21.3**	0.9264	24.4	0.7892	23.9	0.7528	26.2	0.6311	32.4

Table 2. Performance of *SymbolicVision* on multi-tasks.

Input Data	Task	R^2	Nodes
R, Nir	NDVI	0.9733 ± 0.0021	29.2
G, Nir	GNDVI	0.9733 ± 0.0020	47.3
R, Nir	SAVI	0.9675 ± 0.0034	29.1
B, R, Nir	EVI	0.9420 ± 0.0041	37.5
G, Nir	NDWI	0.9975 ± 0.0003	43.1
H	AGB	0.9998 ± 0.0001	33.3
H	CS	0.9995 ± 0.0002	21.8

2. Methodology

Overview. *SymbolicVision* includes a training phase where image features are used to learn symbolic expressions and an inference phase that applies learned expressions to new data. The model is trained with a composite loss.

Symbolic Inference. Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$, where $\mathbf{x}_i$ is a feature vector extracted from imagery, y_i is the target value, and N is the total number of training samples. The goal is to discover an expression $f(\cdot)$ such that $y_i \approx f(\mathbf{x}_i)$. Expressions are constructed from a predefined operator set $\mathcal{O}$, and the search is guided by a Transformer-based decoder.

Encoder Design. We adopt a Swin Transformer (Liu et al., 2021) as the image encoder to extract multi-scale spatial-spectral features $\mathbf{F}_l$. These features are aligned with symbolic regression inputs via a consistency loss $\mathcal{L}_{\text{con}} = \frac{1}{N}\sum_{i=1}^N \|\mathbf{F}_i - \mathbf{F}_i^{\text{target}}\|_2^2$.

Physical Constraints. To ensure physical plausibility, we introduce a regularization term based on energy conservation, formulated as $\mathcal{L}_{\text{phy}} = \sum_k \|\nabla \cdot E_k - \rho_k\|_2^2$, where E_k represents energy flow and ρ_k denotes energy density.

Training Losses. The symbolic decoder is trained with two objectives: a regression loss $\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^N (y_i - f(\mathbf{x}_i))^2$ that ensures numerical accuracy, and a cross-entropy loss $\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^N S_i^{\text{target}} \log(\hat{S}_i)$ that encourages structural similarity between generated and reference expressions.

Module Design. The framework consists of four modules: an image encoder (Swin Transformer) to extract image features $\mathbf{F}_{\text{img}}$; an expression encoder (Transformer) to embed symbolic expressions as $\mathbf{F}_{\text{exp}}$; a decoder that generates expressions using cross-attention; and a feature fusion module that applies multi-head attention to align $\mathbf{F}_{\text{img}}$ and $\mathbf{F}_{\text{exp}}$.

Optimization. The training process consists of three stages. In the first stage, the Swin encoder is pretrained by minimizing the contrastive loss $\mathcal{L}_{\text{con}}$ to align multi-spectral features. In the second stage, the full model is optimized jointly using a combination of losses: contrastive ($\mathcal{L}_{\text{con}}$), cross-entropy ($\mathcal{L}_{\text{CE}}$), mean squared error ($\mathcal{L}_{\text{MSE}}$), and physics-based consistency ($\mathcal{L}_{\text{phy}}$). Finally, in the third stage, the symbolic expressions generated by the model are refined using BFGS optimization. The overall loss function is defined as: $\mathcal{L} = \lambda_{\text{con}}\mathcal{L}_{\text{con}} + \lambda_{\text{CE}}\mathcal{L}_{\text{CE}} + \lambda_{\text{MSE}}\mathcal{L}_{\text{MSE}} + \lambda_{\text{phy}}\mathcal{L}_{\text{phy}}$.

3. Experiments

We evaluate *SymbolicVision* on three datasets: SRBench (de Franca et al., 2024), Open-Canopy (Fogel et al., 2024), and an Expanded Open-Canopy dataset (Appendix A). For evaluation, we adopt R^2, and Nodes to quantify. Experiments are conducted on an NVIDIA A100 GPU.

Comparison with baselines. Table 1 shows the performance of *SymbolicVision* against recent SR baselines including MMSR, TPSR, SymbolicGPT, and NeSymReS. Our method consistently achieves top-2 accuracy while maintaining fewer symbolic nodes, reflecting a favorable trade-off between interpretability and performance.

Multi-task symbolic regression. Table 2 demonstrates *SymbolicVision*'s capability on multi-modal remote sensing tasks. It achieves R^2 above 0.97 on vegetation indices such as NDVI and GNDVI, and over 0.999 on biomass (AGB) and carbon stock (CS), verifying strong generalization in symbolic prediction.

4. Conclusion and Future Work

SymbolicVision enables symbolic regression directly from multi-spectral remote sensing imagery by combining Swin-based feature extraction, a symbolic Transformer decoder, and physics-aware constraints. It achieves both high accuracy and interpretability across geospatial tasks. Future work includes extending to hyperspectral data and integrating uncertainty estimation and symbolic priors.

References

AbdusSalam, S., Abel, S., and Romão, M. C. Symbolic regression for beyond the standard model physics. *Physical Review D*, 111(1):015022, 2025.

Biggio, L., Bendinelli, T., Neitz, A., Lucchi, A., and Parascandolo, G. Neural symbolic regression that scales. *International Conference on Machine Learning (ICML)*, pp. 936–945, 2021. Placeholder; replace with actual publication details.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology*, 20(10):3177–3190, 2014.

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016. doi: 10.1145/2939672.2939785.

Davis, E. Mathematics, word problems, common sense, and artificial intelligence. *Bulletin of the American Mathematical Society*, 61(2):287–303, 2024.

de Franca, F. O., Virgolin, M., Kommenda, M., Majumder, M., Cranmer, M., Espada, G., Ingelse, L., Fonseca, A., Landajuela, M., Petersen, B., et al. Srbench++: Principled benchmarking of symbolic regression with domain-expert interpretation. *IEEE transactions on evolutionary computation*, 2024.

Fogel, F., Perron, Y., Besic, N., Saint-André, L., Pellissier-Tanon, A., Schwartz, M., Boudras, T., Fayad, I., d'Aspremont, A., Landrieu, L., et al. Open-canopy: A country-scale benchmark for canopy height estimation at very high resolution. *arXiv preprint arXiv:2407.09392*, 2024.

Gitelson, A. A. and Merzlyak, M. N. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *Journal of plant physiology*, 148(3-4):494–500, 1996.

Glazer, E., Erdil, E., Besiroglu, T., Chicharro, D., Chen, E., Gunning, A., Olsson, C. F., Denain, J.-S., Ho, A., Santos, E. d. O., et al. Frontiermath: A benchmark for evaluating advanced mathematical reasoning in ai. *arXiv preprint arXiv:2411.04872*, 2024.

Gong, Z., Ge, W., Guo, J., and Liu, J. Satellite remote sensing of vegetation phenology: Progress, challenges, and opportunities. *ISPRS Journal of Photogrammetry and Remote Sensing*, 217:149–164, 2024.

Grayeli, A., Sehgal, A., Costilla Reyes, O., Cranmer, M., and Chaudhuri, S. Symbolic regression with a learned concept library. *Advances in Neural Information Processing Systems*, 37:44678–44709, 2024.

Hochmair, H. H., Juhász, L., and Li, H. Advancing ai-driven geospatial analysis and data generation: Methods, applications and future directions, 2025.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote sensing of environment*, 83(1-2):195–213, 2002.

Huete, A., Justice, C., and Liu, H. Development of vegetation and soil indices for remote sensing. *Remote Sensing of Environment*, 216:123–135, 2018. doi: 10.1016/j.rse.2018.06.027. Placeholder for NDVI/SAVI; replace with specific reference.

Huete, A. R. A soil-adjusted vegetation index (savi). *Remote sensing of environment*, 25(3):295–309, 1988.

Jiang, D., Li, Y., Liu, Q., and Huang, C. Evaluating the sustainable development science satellite 1 (sdgsat-1) multi-spectral data for river water mapping: A comparative study with sentinel-2. *Remote Sensing*, 16(15):2716, 2024.

Kamienny, P.-A., d'Ascoli, S., Lample, G., and Charton, F. End-to-end symbolic regression with transformers. *arXiv preprint arXiv:2110.10246*, 2021. Placeholder; replace with actual publication if available.

Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1994. doi: 10.7551/mitpress/1109.001.0001.

Landajuela, M., Petersen, B. K., Glatt, R., and Mundhenk, T. N. Transformer-based planning for symbolic regression. *arXiv preprint arXiv:2306.01723*, 2023. Placeholder; replace with actual publication if available.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the*

IEEE/CVF international conference on computer vision, pp. 10012–10022, 2021.

Makke, N. and Chawla, S. Interpretable scientific discovery with symbolic regression: a review. *Artificial Intelligence Review*, 57(1):2, 2024a.

Makke, N. and Chawla, S. Symbolic regression: A pathway to interpretability towards automated scientific discovery. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 6588–6596, 2024b.

Malerba, M. E., de Kluyver, T., Wright, N., Omosalewa, O., and Macreadie, P. I. Including methane emissions from agricultural ponds in national greenhouse gas inventories. *Environmental Science & Technology*, 58(19):8349–8359, 2024.

McFeeters, S. K. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International journal of remote sensing*, 17(7):1425–1432, 1996.

Merler, M., Haitsiukevich, K., Dainese, N., and Marttinen, P. In-context symbolic regression: Leveraging large language models for function discovery. *arXiv preprint arXiv:2404.19094*, 2024.

Polydoros, G., Galitskaya, V., Pergantis, P., Drigas, A., Antoniou, A.-S., and Beazidou, E. Innovative ai-driven approaches to mitigate math anxiety and enhance resilience among students with persistently low performance in mathematics. *Psychology International*, 7(2):46, 2025.

Raj, T., Verma, S., Kumar, N., and Agrawal, R. Role of woody biomass in carbon capture, circular bioeconomy, and biomanufacturing. In *Sustainable Biorefining of Woody Biomass to Biofuels and Biochemicals*, pp. 291–318. Elsevier, 2024.

Tashtoush, M. A., Wardat, Y., Ali, R. A., and Saleh, S. Artificial intelligence in education: mathematics teachers' perspectives, practices and challenges. *Iraqi Journal for Computer Science and Mathematics*, 5(1):20, 2024.

Tian, Y., Zhou, W., Viscione, M., Dong, H., Kammer, D. S., and Fink, O. Interactive symbolic regression with co-design mechanism through offline reinforcement learning. *Nature Communications*, 16(1):3930, 2025.

Udrescu, S.-M. and Tegmark, M. Ai feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020. doi: 10.1126/sciadv.aay2631.

Wang, G., Wang, E., Li, Z., Zhou, J., and Sun, Z. Exploring the mathematic equations behind the materials science data using interpretable symbolic regression. *Interdisciplinary Materials*, 3(5):637–657, 2024.

Yang, K., Poesia, G., He, J., Li, W., Lauter, K., Chaudhuri, S., and Song, D. Formal mathematical reasoning: A new frontier in ai. *arXiv preprint arXiv:2412.16075*, 2024.

Yu, Z. Ai for science: A comprehensive review on innovations, challenges, and future directions. *International Journal of Artificial Intelligence for Science (IJAI4S)*, 1 (1), 2025.

Yu, Z., Wang, J., and Idris, M. Y. I. Iidm: Improved implicit diffusion model with knowledge distillation to estimate the spatial distribution density of carbon stock in remote sensing imagery. *arXiv preprint arXiv:2411.17973*, 2024.

Yu, Z., Chen, H., Idris, M. Y. I., and Wang, P. Rainy: Unlocking satellite calibration for deep learning in precipitation. *arXiv preprint arXiv:2504.10776*, 2025a.

Yu, Z., Idris, M., and Wang, P. Satellitecalculator: A multi-task vision foundation model for quantitative remote sensing inversion. *arXiv preprint arXiv:2504.13442*, 2025b.

Yu, Z., Idris, M. Y. I., Wang, H., Wang, P., Chen, J., and Wang, K. From physics to foundation models: A review of ai-driven quantitative remote sensing inversion. *arXiv preprint arXiv:2507.09081*, 2025c.

Yu, Z., Idris, M. Y. I., and Wang, P. Dc4cr: When cloud removal meets diffusion control in remote sensing. *arXiv preprint arXiv:2504.14785*, 2025d.

Yu, Z., Idris, M. Y. I., and Wang, P. Satellitemaker: A diffusion-based framework for terrain-aware remote sensing image reconstruction. *arXiv preprint arXiv:2504.12112*, 2025e.

Zhang, P., Tang, K., Wang, A., Wu, H., and Zhong, Z. Neural network integrated with symbolic regression for multiaxial fatigue life prediction. *International Journal of Fatigue*, 188:108535, 2024a.

Zhang, Y., Li, X., Hu, W., and Yen, G. G. Multiexpression symbolic regression and its circuit design case. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024b.

Zhou, R., Xu, M., Chen, S., Liu, J., Li, Y., Xinxin, L., Chen, Z., and He, J. Math for ai: On the generalization of learning mathematical problem solving. 2024.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017. doi: 10.1109/MGRS.2017.2762307.

A. Constructing Dataset

A.1. Parameter Description

The spectral indices used in this study are derived from **SPOT 6/7** data and serve as key indicators for vegetation condition and ecological structure estimation. All computations are based on four spectral bands: **Blue** (**B1**, 450–520 nm), **Green** (**B2**, 530–590 nm), **Red** (**B3**, 625–695 nm), and **Near Infrared** (**B4**, 760–890 nm).

We include five vegetation-related indices: NDVI, GNDVI, SAVI, EVI, and NDWI. Among them, SAVI introduces a soil adjustment factor L to correct background influence, while EVI incorporates gain and aerosol resistance coefficients G, C_1, and C_2 for atmospheric correction. These indices are calculated using standard definitions widely adopted in remote sensing literature.

In addition to indices, we include three structural variables: canopy height (H), aboveground biomass (AGB), and carbon stock (CS), which are estimated from regression models using NDVI, SAVI, and LiDAR-derived height information. Empirical coefficients (a, b, c, CF) in these models are region-specific and derived from ecological studies on open-canopy forest systems.

A.2. Spectral Indices

1. **Normalized Difference Vegetation Index (NDVI)** (Gong et al., 2024)

$$NDVI = \frac{B4 - B3}{B4 + B3} \tag{1}$$

   NDVI is widely used to measure vegetation health and biomass productivity. It utilizes the difference between near-infrared (NIR, B4) and red (B3) reflectance to assess chlorophyll activity in plants.

2. **Green Normalized Difference Vegetation Index (GNDVI)** (Gitelson & Merzlyak, 1996)

$$GNDVI = \frac{B4 - B2}{B4 + B2} \tag{2}$$

   GNDVI is a modification of NDVI that enhances sensitivity to vegetation chlorophyll content by incorporating the green band (B2) instead of the red band (B3).

3. **Soil-Adjusted Vegetation Index (SAVI)** (Huete, 1988)

$$SAVI = \frac{(B4 - B3) \times (1 + L)}{B4 + B3 + L} \tag{3}$$

   where $L = 0.5$ is the soil adjustment factor, which minimizes soil brightness effects in low-vegetation areas.

4. **Enhanced Vegetation Index (EVI)** (Huete et al., 2002)

$$EVI = G \times \frac{(B4 - B3)}{B4 + C_1 \times B3 - C_2 \times B1 + L} \tag{4}$$

   where $G = 2.5$, $C_1 = 6$, $C_2 = 7.5$, and $L = 1$. EVI improves upon NDVI by reducing atmospheric and background noise, making it more suitable for dense vegetation monitoring.

5. **Normalized Difference Water Index (NDWI)** (McFeeters, 1996)

$$NDWI = \frac{B2 - B4}{B2 + B4} \tag{5}$$

   NDWI is used to monitor water bodies, distinguishing open water from land features. It leverages the high reflectance of water in the green band (B2) and the strong absorption in the NIR band (B4).

A.3. Ecological Structural Variables

1. **Canopy Height (H)** The canopy height data is directly obtained from the Open-Canopy dataset (Fogel et al., 2024) and is not derived from NDVI or any regression model.

2. **Aboveground Biomass (AGB)** (Chave et al., 2014)

$$AGB = a \times H^b \tag{6}$$

where H represents canopy height obtained from the Open-Canopy dataset, and a, b are empirical constants derived from literature or site-specific calibration.

For temperate forests in France, typical values from literature suggest that for coniferous forests, the empirical coefficients are $a = 0.118$ and $b = 2.53$; for broadleaf forests, $a = 0.052$ and $b = 2.69$; and for mixed forests, $a = 0.067$ and $b = 2.58$ (Raj et al., 2024).

If forest type is unknown, a general model can be used:

$$AGB = 0.067 \times H^{2.58} \tag{7}$$

which is a widely used allometric equation for estimating biomass in temperate forests.

3. **Carbon Stock (CS)** (Malerba et al., 2024)

$$CS = AGB \times CF \tag{8}$$

where CS represents the carbon stock (tC/ha), which is the total amount of carbon stored in aboveground biomass. AGB refers to the aboveground biomass (t/ha), estimated from canopy height using empirical models. $CF = 0.47$ is the carbon fraction factor, indicating that approximately 47% of the total biomass is composed of carbon, as suggested by IPCC guidelines. This factor may vary depending on tree species and environmental conditions but is commonly used in temperate and tropical forest studies.