Contextualizing Pan-Tropical Allometric Models for Biomass Estimation

Eustache Diemert, Anaëlle Dambreville

PUR

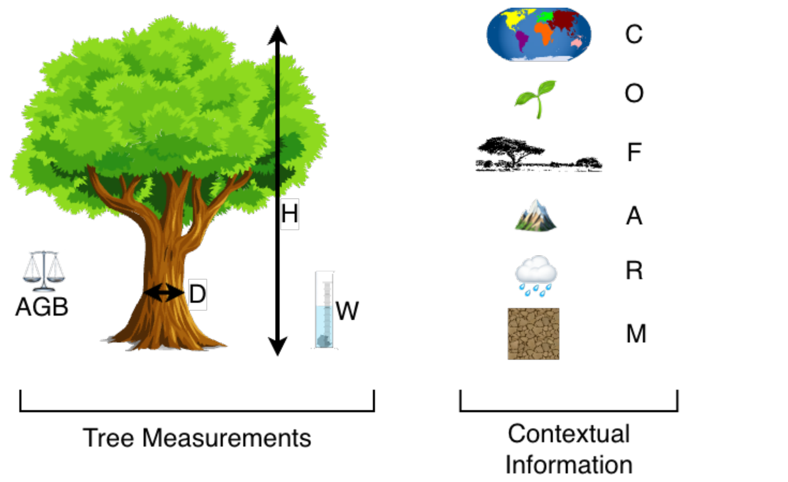
Introduction

Allometric Models (AMs) play a central role in monitoring and mitigating climate change as they provide accurate estimation of biomass and carbon sequestered by trees from non-destructive, easy to obtain physical measurements.

Unfortunately, practitioners spend considerable effort in researching, qualifying and choosing AMs for specific growth conditions.

Classical Pan-Tropical Allometric Models

The dataset from **(author?)** [1] contains measurements of trunk diameter at breast height D (cm), total tree height H (m), wood density W ($g.cm^{-3}$), and total oven-dry above ground dry matter AGB (kg) for 4004 trees across 58 sites worldwide.



	Tree Measurements
D	diameter at breast height (cm)
H	total height (m)
W	wood density ($g.cm^{-3}$)
	Contextual Information
C	continent ($categorical$)
O	old growth ($binary$)
F	forest type ($categorical$)
A	altitude (m)
R	annual rainfall (mm)
M	dry months ($0-12$)
	Tree Biomass (prediction target)
AGB	above ground dry matter (kg)

Figure 1. Tree measurements and contextual variables

The state of the art pan-tropical allometric model from [1]: $AGB = \alpha(D^2HW)^\beta$

Frontiers in Allometric Models Development

- Can allometric models be improved by incorporating richer contextual information ?
- Can transfer error be predicted when using an allometric model in a new site ?

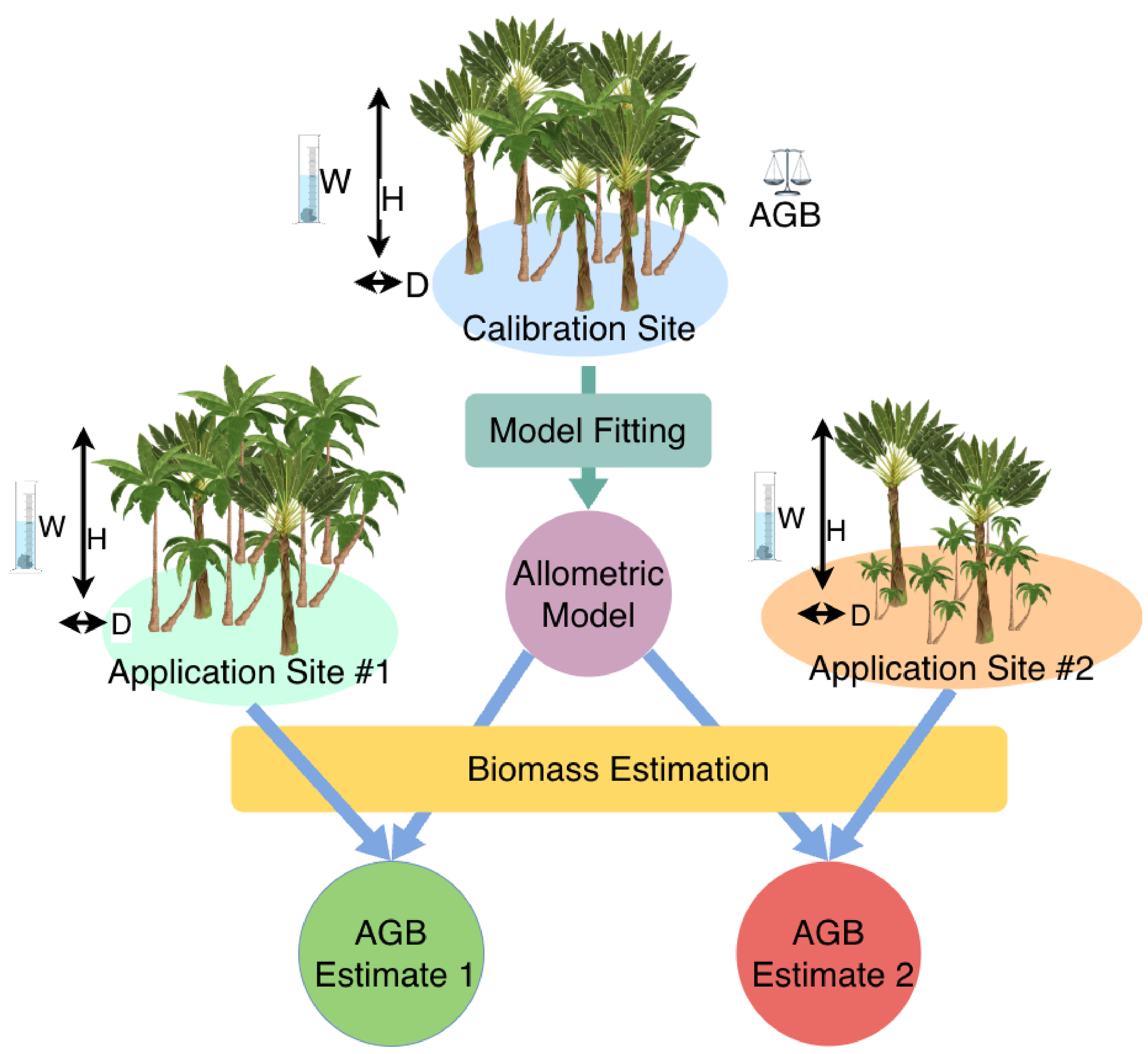


Figure 2. Illustration of how data distribution shift between Calibration site and two Application sites may affect estimated Biomass: a severe shift at site 2 jeopardizes estimated AGB Accuracy while a mild shift doesn't affect site 1 accuracy.

Contribution #1: Deep Allometric Model

First, [1] can be trained as $logAGB = \alpha logD + \beta logH + \gamma logW$ with L^2 regularization. Then, assuming contextual information can provide valuable insights to the model, we add *ContextualChave*:

$$AGB = \alpha(D^2HW)^{\beta_0}\tilde{C}^{\beta_1}\tilde{O}^{\beta_2}\tilde{F}^{\beta_3}A^{\beta_4}R^{\beta_5}M^{\beta_6} \quad (1)$$

where $\tilde{}$ designates Target Encoding, a popular and efficient method to encode categorical variables.

Finally, we propose deep learning models:

- *COFARM* a model that embeds categorical covariates C, O, F in 2 dimensions before pooling them with D, H, W, A, R, M in a log-log linear regressor
- *COFARM-NN*, inspired by wide and deep architectures pools embedded categorical covariates C, O, F with A, R, M numerical features and passes resulting information through 3 layers of 3 fully connected neurons and ReLu blocks before the final regressor, as depicted on Figure 3.

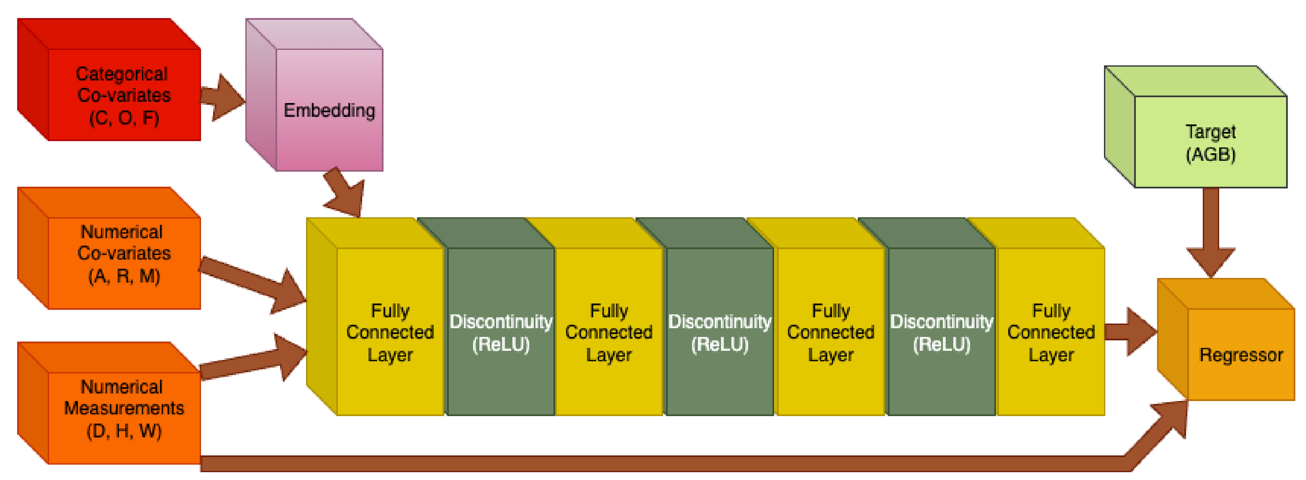


Figure 3. *COFARM-NN* model architecture for AGB prediction

Contribution #2: Transfer Error Model

We rely on [2] that establish a theoretical upper bound on the expected target error $\epsilon_t(f)$:

$$\epsilon_t(h) \leq \epsilon_s(h) + \eta_{\mathcal{H}}(f_s, f_t) + \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h) \quad (2)$$

where $\eta_{\mathcal{H}}(f_s, f_t)$ is the sum of the errors made by the ideal predictor on both source and target domains and $\mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h)$ is the hypothesis-discrepancy associated with h and defined as

$$\mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h) = \max_{h' \in \mathcal{H}} |\mathbb{E}_{x\sim p_t}[L(h'(x), h(x))] - \mathbb{E}_{x\sim p_s}[L(h'(x), h(x))]| \quad (3)$$

While η is purely related to the labeling discrepancy, $\mathcal{H}Disc$ measures the discrepancy between source and target distributions, relatively to the model h.

We propose a quadratic model of the additional prediction error:

$$\Delta_{s,t}MAE(h) = \alpha \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h) + \beta \times sign(\Delta_{s,t}Z) + \gamma \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h)^2 + \delta \times (sign(\Delta_{s,t}Z) \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h)) \quad (4)$$

where $\Delta_{s,t}Z = \mathbb{E}_{X\sim p_s}[D^2HW] - \mathbb{E}_{X\sim p_t}[D^2HW]$. The model is learned as a Ridge regression with L^2 penalty strength optimized by internal cross-validation.

To estimate $\mathcal{H}Disc$ in practice we solve the following constrained optimization problem

$$\max_{\alpha',\beta'} |(\hat{\epsilon}_s(f_{\alpha,\beta}) - \hat{\epsilon}_s(f_{\alpha',\beta'})) - (\hat{\epsilon}_t(f_{\alpha,\beta}) - \hat{\epsilon}_t(f_{\alpha',\beta'}))| \text{ s.t. } |\alpha' - \alpha| < \delta_1 \text{ and } |\beta' - \beta| < \delta_2 \quad (5)$$

, where $\hat{\epsilon}$ is the empirical MAE.

Experiment: Deep Allometric Models Performance

We benchmark a wide range of models to understand what can lead to reduced prediction errors. Baseline: [1] and simple log-log regression (*LogReg*); More complex: MLP (*LogReg-NN*); More features: Perceptron + Embeddings (*COFARM*); More complex and more features: tree based (*HGBRT*), MLP + Embeddings (*COFARM-NN*).

Table 1. Predictive power of different AMs (**bold** = best average, * = stat. sig. at 90% vs [1])

Dataset	Metric	[1]	*ContextualChave*	*HGBRT*	*LogReg*	*LogReg-NN*	*COFARM*	*COFARM-NN*
train	RMSE	829.9 ± 23.3	822.8 ± 21.2	628.8 ± 24.8	762.3 ± 18.9	762.3 ± 18.9	692.2 ± 20.2	**590.0** ± 35.2
train	MAE	175.0 ± 2.9	163.4 ± 2.3	**83.2** ± 2.4	168.8 ± 2.6	168.8 ± 2.6	147.7 ± 2.3	130.6 ± 3.8
train	R^2	0.888 ± 0.007	0.890 ± 0.006	0.936 ± 0.004	0.905 ± 0.006	0.905 ± 0.006	0.922 ± 0.005	**0.942** ± 0.007
test	RMSE	804.9 ± 80.7	797.6 ± 78.9	785.4 ± 111.1	747.3 ± 69.8	747.2 ± 69.8	**694.6** ± 71.7	702.9 ± 91.7
test	MAE	180.4 ± 11.1	168.0 ± 9.3	170.5 ± 13.7	173.6 ± 10.2	173.6 ± 10.2	155.4 ± 9.2* [-14%]	**148.9** ± 9.4* [-17%]
test	R^2	0.859 ± 0.025	0.866 ± 0.027	0.813 ± 0.062	0.865 ± 0.030	0.865 ± 0.030	0.880 ± 0.029	**0.881** ± 0.026

Experiment: Predict Transfer Error

We experiment with synthetic covariate shifts by splitting the data according to explanatory variables. Then the model from [1] is learned on one split and tested and the other. Finally we compare the additional transfer error $\Delta_{s,t}MAE(h)$ as predicted by Eq. 4 vs as estimated from observed empirical error.

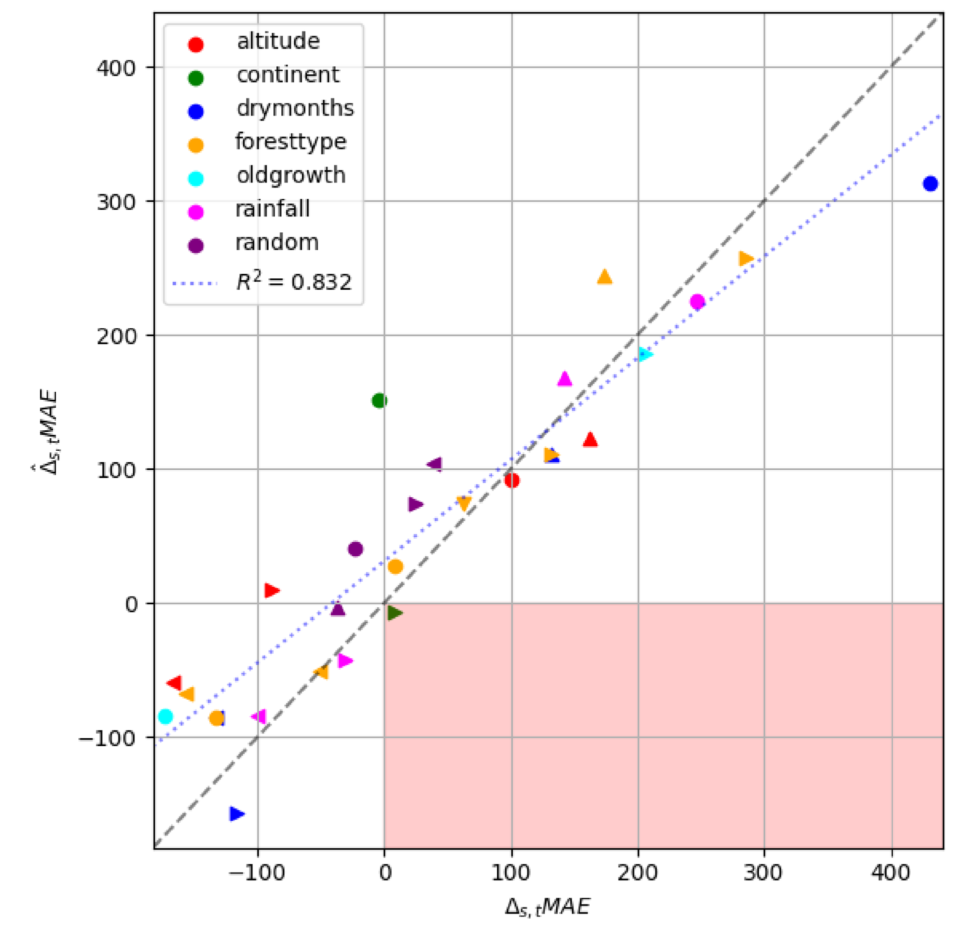


Figure 4. **Predicted vs Actual Additional Error due to Covariate Shift**. Color indicates the variable chosen to simulate the shift while marker shapes indicate different shifts magnitudes and directions.

Conclusion

- Embedding categorical contextual information leads to -14% MAE, while adding deep layers reduces further MAE by -3%
- Additional transfer error when applying a model to a new plot *can be predicted* ($R^2 = .832$) by quantifying the discrepancy between source and target features distribution

References

[1] J. Chave, M. Réjou-Méchain, A. Búrquez, E. Chidumayo, M. S. Colgan, W. B. Delitti, A. Duque, T. Eid, P. M. Fearnside, R. C. Goodman, M. Henry, A. Martínez-Yrízar, W. A. Mugasha, H. C. Muller-Landau, M. Mencuccini, B. W. Nelson, A. Ngomanda, E. M. Nogueira, E. Ortiz-Malavassi, R. Pélissier, P. Ploton, C. M. Ryan, J. G. Saldarriaga, and G. Vieilledent. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20(10):3177–3190, Oct. 2014.

[2] G. Richard, A. de Mathelin, G. Hébrail, M. Mougeot, and N. Vayatis. Unsupervised Multi-Source Domain Adaptation for Regression. In *Lecture Notes in Computer Science book series (LNCS, volume 12457)*, volume 12457 of *Lecture Notes in Computer Science book series*, pages 395–411, Ghent, Belgium, Sept. 2020. Springer International Publishing.

eustache.diemert@pur.co