Contextualizing Pan-Tropical Allometric Models for Biomass Estimation

Eustache Diemert[1] **and Anaëlle Dambreville**[1]

[1]PUR Data Team, The PUR Projet, France

Correspondence: Eustache Diemert (eustache.diemert@pur.co)

Abstract. 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. To overcome this situation Chave et al. (2014) developed a pan-tropical AM with equivalent accuracy to local, site-specific AMs. We ameliorate this result by incorporating contextual information pertaining to growth conditions in a Machine Learning (ML) model, eventually achieving a reduction in Mean Average Error (MAE) of -17% as measured on hold-out data. This breakthrough shall have important impact in applications such as national forest inventories, carbon certifications and calibration of satellite based biomass maps to field data. To complete, we propose a principled method to estimate how much additional error one can expect when applying a given AM to shifting conditions and provide a data-driven safety check to practitioners.

1 Introduction

AMs support numerous important applications such as National Forest Inventories (NFIs), local afforestation and reforestation projects driven by carbon markets and large-scale scientific efforts to monitor the Earth system. In typical instances AMs are used directly to convert trees measurements in the field into biomass. Recently, this has been combined with Earth Observation (EO) systems that provide Above-Ground Biomass (AGB) maps. These maps rely indirectly on AMs as AGB models predicting AGB from air or space-borne canopy measurements are calibrated to AGB estimates provided by NFIs (Hunka et al., 2024; Santoro and Cartus, 2025). As such, even small improvement in AMs estimation accuracy can lead to potentially huge corrections, as the scale of application encompasses millions of tons of biomass and sequestered carbon (Blaufelder et al.).

Tree allometry has a long history, with early efforts dating from the work of Smith and Brand (1983). It has grown as a research field as modern computing and data became more and more accessible. Nowadays AMs are available for a variety of conditions affecting tree biomass including genetic (e.g. tree species), environmental (e.g. temperature, light, soil) and management-related (e.g. agroforestry system). Efforts have been made to consolidate them in general purpose databases (Henry et al., 2013) or more focused reviews, some of them encompassing several hundreds equations for a single country (Rojas-García et al., 2015). At the same time the scope of application of AMs expands, leading scientists to question how to properly select an AM for specific growth conditions (Van Breugel et al., 2011) as different AMs lead sometimes to very different estimations (Zhao et al., 2012). In practice, scientists and engineers select AMs developed in the closest possible conditions (Lovell and MacKenzie, 2014), as developing new AMs incur cutting, drying and weighing at least 40 trees, a cost that few projects can afford (Návar, 2010). However, there is to the best of our knowledge no quantitative method that can evaluate if one can safely use an AM developed for another context.

At this point we'd like to draw attention to recent developments that highlight an opportunity to improve applicability and accuracy of AMs. First, researchers have started exploring the use of ML models to replace classical power models and reported success in regional studies (Dutta Roy and Debbarma, 2024; Wongchai et al., 2022). Second, a pan-tropical AM introduced in (Chave et al., 2005) and enhanced in (Chave et al., 2014) has been showed to successfully replace site-specific AMs by a single model learned on pooled data from many sites encompassing different continents, for-

Table 1. Problem Variables

	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)

est types and species. Their model only relies on 3 easily accessible parameters: diameter, height and wood density of trees. Considering AM development from a Statistical Learning point of view, we point out that prediction accuracy on unseen data is the result of a trade-off between data size and model complexity (Luxburg and Schoelkopf, 2008). The pan-tropical AM from Chave et al. (2014) pools data from 58 sites and learns a 3 parameter power model. Following this line of thought we postulate that incorporating additional *contextual information* in the model would grow complexity yet has a chance to enhance prediction accuracy by better describing factors affecting biomass production. This idea is illustrated in Figure 1.

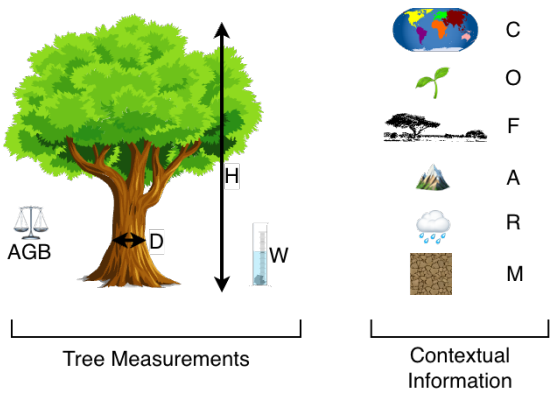

Figure 1. Tree measurements and contextual variables in the biomass estimation problem. Variables described in Table 1.

On the question of choice between available AMs, another insight from statistical learning is that the shift in data distributions between the calibration site (used to fit model coefficients) and application site (where inference takes place) might derail generalization of a model and produce inaccurate estimates. This situation is illustrated in Figure 2 and commonly known in statistical learning as Domain Adaptation (DA) (Cortes and Mohri, 2014). Here, we propose to use distribution distance metrics to quantify the shift (Ben-David et al., 2010; Richard et al., 2020) between training and application sites and learn a simple approximation of the related biomass prediction error.

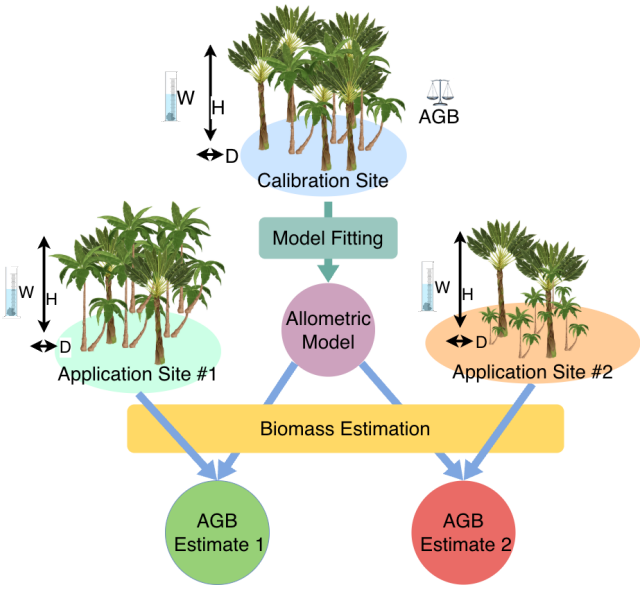

Figure 2. Illustration of how data distribution shift between Calibration site and two Application sites may affect estimated Biomass: a severe shift at site 2 jeopardizes estimated AGB Accuracy while a mild shift doesn't affect site 1 accuracy.

Our contributions can be summarized as follows:

1. *COFARM-NN*, a deep learning model achieving a reduction in MAE of -17% vs Chave et al. (2014)

2. *DiscShift*, a practical, data-driven method to estimate additional error when using an AM on new application sites

2 Material and methods

2.1 Data

We use the dataset from Chave et al. (2014) as material for all of our experiments. It encompasses 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. To learn contextual models we use additional covariates provided in (Chave et al., 2005) for a portion of the sites and described in Table 1. We also exclude sites that were found to contain unreliable data and later removed in Chave et al. (2014) to obtain a contextual dataset of 1,481 trees.

2.2 AM Models

2.2.1 Models definition

We compare different models against Model 1 from Chave et al. (2014):

$$AGB = \alpha(D^2HW)^\beta \tag{1}$$

trained as explained in Eq. 2 of the original paper.

First, we select context-agnostic baselines. We train *LogReg*, a simple log-log linear regression of

$$logAGB = \alpha logD + \beta logH + \gamma logW \tag{2}$$

to which we adjunct an L^2 regularization. We also train *HGBRT*, a Gradient Boosted Decision Trees regression model with a maximum of 1,000 iterations, access to all covariates (tree measurements & contextual information) (Shi et al., 2022) and hyper-parameter optimization to find relevant regularization parameters.

Then, assuming contextual information can provide valuable insights to the model, we add *ContextualChave*:

$$AGB = \alpha(D^2HW)^{\beta_0}\tilde{C}^{\beta_1}\tilde{O}^{\beta_2}\tilde{F}^{\beta_3}A^{\beta_4}R^{\beta_5}M^{\beta_6} \tag{3}$$

where $\tilde{}$ designates Target Encoding (Pargent et al., 2022), a popular and efficient method to encode categorical variables.

Finally, we propose 3 deep learning models:

- *LogReg-NN*, an evolution of the log-log regression on D, H, W with 2 additional fully connected layers of 3 neurons before the final regressor
- *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 (Cheng et al., 2016) 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.

These models are implemented in PyTorch (Paszke et al., 2019) and trained with the Adam optimizer (Kingma and Ba, 2017).

2.2.2 Models evaluation

We perform 30 random train/test splits of proportion .8/.2 on the dataset of Chave et al. (2014), learning each model on train data and evaluating on hold-out test data. We report average value over the 30 splits and 90%-tile confidence intervals (Sileshi, 2014). Note that this methodology incorporates randomness of both data distribution and learning for non-convex methods such as deep neural networks.

2.3 Additional Error Prediction in Shifting Conditions

2.3.1 Preliminary: Unsupervised Domain Adaptation

The application of AMs to new sites can be seen as a problem of Unsupervised Domain Adaptation (DA). Indeed, coefficients of an AM fitted on a calibration site depend on the distribution of covariates. These coefficients may be shifted compared to the application site because of the specific growth conditions of the application site compared to the calibration one. For instance, this can happen when an AM developed for a given biome and species mix in a primary forest is applied to plots from a similar biome and species mix but in a secondary forest. It is likely in this case that the AM will make a larger error in predicting the target AGB as its assumptions would be violated to a certain point when applying it on data from the application site. In the remaining of the paper, to be consistent with DA vocabulary we shall call *source* distribution the distribution of covariates in the calibration site and *target* distribution the one of the application site.

In statistical terms, we consider a predictor $h : \mathcal{X} \to \mathcal{Y}$, source and target datasets $\mathcal{D}_s = (X_s, f_s)$ and $\mathcal{D}_t = (X_t, f_t)$ and a loss function $L : (\mathcal{Y}, \mathcal{Y}) \to \mathbb{R}$. We are interested in the expected error in the target domain of a hypothesis (or predictor) $h \in \mathcal{H}$: $\epsilon_t(h) = \mathbb{E}_{x\sim p_t}[L(f_t, h(x))]$. To develop a practical estimator $\hat{h}$ we assume access to an empirical source dataset $\hat{\mathcal{D}}_s = \{x_s, f_s\}_{i=1\ldots n}$ and target covariates $\hat{X}_t$; that allows to estimate the expected error of h on source $\epsilon_s(h) = \mathbb{E}_{x\sim p_s}[L(f_s, h(x))]$ while the true target label f_t remains unobserved (Wilson and Cook, 2020).

Many approaches have been proposed to tackle this problem, mostly in the classification case and applied to image related tasks (Gulrajani and Lopez-Paz, 2020). For regression, we rely on (Richard et al., 2020) that establish a theoretical upper bound on the expected target error $\epsilon_t(f)$ using tools from statistical learning theory. We restate their Proposition 1 for convenience:

$$\epsilon_t(h) \leq \epsilon_s(h) + \eta_{\mathcal{H}}(f_s, f_t) + \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h) \tag{4}$$

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

$$\mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h) = \max_{h'\in\mathcal{H}} |\mathbb{E}_{x\sim p_t}[L(h'(x), h(x))] - \mathbb{E}_{x\sim p_s}[L(h'(x), h(x))]| \tag{5}$$

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. In other words, this bound suggests that the additional error made when applying a regression model learned on a source dataset to a tar-

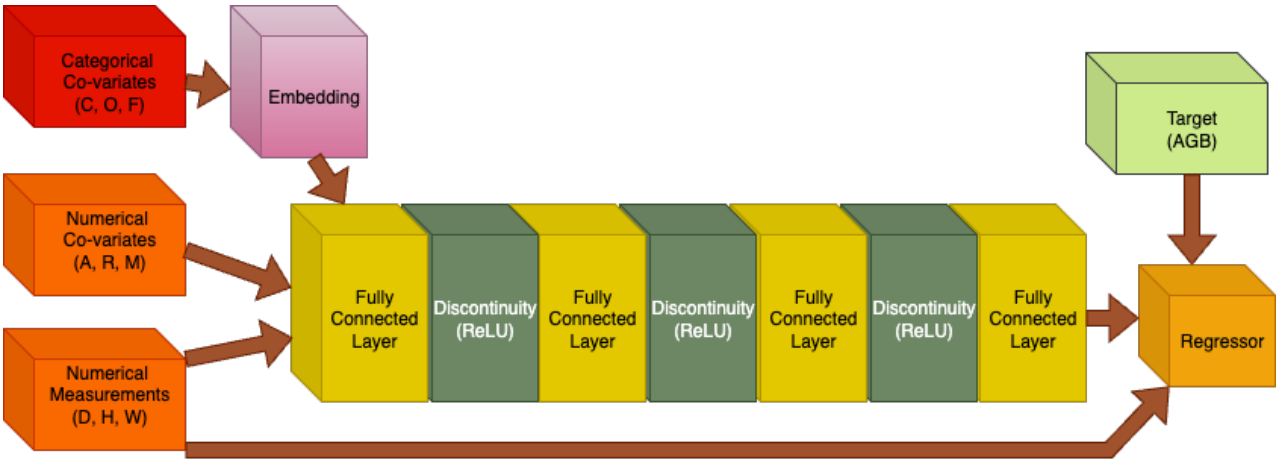

Figure 3. *COFARM-NN* model architecture for AGB prediction. Refer to Table 1 for description of variables.

get dataset can be quantified by measuring the shift between source and target covariates distributions.

2.3.2 Additional Error Prediction Model

Considering that the prediction error in the target domain can be related to the shift between source and target covariates, we postulate that the suitability of an AM to new, potentially shifting conditions can be modeled. Let us consider $\Delta_{s,t} = \epsilon_t(h) - \epsilon_s(h)$, the additional error made when applying an AM to new, shifted conditions. We consider from now on L to be the MAE which is both telling for practical usage and respects the triangle inequality and symmetry assumptions in Richard et al. (2020). We then propose a degree 2 polynomial model of the additional error

$$\begin{aligned} \Delta_{s,t} MAE(h) &= \alpha \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h) \\ +\beta \times sign(\Delta_{s,t} Z) &+ \gamma \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h)^2 \\ +\delta \times (sign(\Delta_{s,t} Z) &\times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h)) \end{aligned} \quad (6)$$

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

$$\max_{\alpha', \beta'} \left| \left(\hat{\epsilon}_s(f_{\alpha,\beta}) - \hat{\epsilon}_s(f_{\alpha',\beta'})\right) - \left(\hat{\epsilon}_t(f_{\alpha,\beta}) - \hat{\epsilon}_t(f_{\alpha',\beta'})\right) \right| \quad \text{s.t. } |\alpha' - \alpha| < \delta_1 \text{ and } |\beta' - \beta| < \delta_2 \quad (7)$$

, where $\hat{\epsilon}$ is the empirical MAE. In other words we assume $\mathcal{H}$ to be the family of models which parameters reside inside a ball of fixed radius centered on coefficients of the Chave model fitted on the source site data. The solution is found by solving successive quadratic optimization with the *SLS-QP* algorithm (Nocedal and Wright, 2006).

2.3.3 Experimental Distribution Shift

We create distribution shifts experimentally in the Chave dataset by splitting it according to contextual information to produce fake calibration and application sites. This ensures existence of a shift in tree measurement and biomass distributions between the two sub-parts of the data. For each numerical variable (A, R, M), we compute the 25th and 75th percentiles and use them as thresholds to divide the distribution into two sub-distributions, corresponding to source and target datasets. Table 2 shows the thresholds used to implement the shifts. For example, for altitude, the 25th percentile is 42.0 m ("altitude-low" shift). The source dataset is thus composed of trees located < 42 m, and the target dataset encompasses the trees located >= 42 m. C (Asia or Americas) and O (0 or 1) have only two categories. Source and target datasets are thus created by separating the trees based on one of the two values. F encompasses four categories. Each one of them has been used to create a supplementary shift in distributions. We also consider two random shifts by randomly separating the dataset, resulting to a total of 14 shift types. For each of them, the objective is to train Chave et al. (2014) AM on the source dataset and to predict AGB and the additional error on the target dataset. We also implement reversed shifts (indicated by $\longleftarrow$) by switching source and target datasets leading to a total of 28 shifts.

To illustrate the magnitude of the discrepancy between source and target datasets for each shift we compute the Wasserstein distance on D^2HW (Panaretos and Zemel (2019)). Moreover, a significant shift in distributions implies that it would be possible to predict the origin of a tree (source or target distribution) based on its covariates values (D^2, H, and W). To illustrate this statement, we train a Gradient Boosted Decision Trees classifier model (Friedman, 2002) on the Chave dataset to predict whether a tree belongs to the source dataset or not, for each shift ((Lopez-Paz and Oquab,

Table 2. Reference values used to experimentally create the shifts with Wasserstein distance and $\mathcal{H}Disc$ between source/calibration and target/application.

Var.	Shift	Reference	Wass. Dist	$\mathcal{H}Disc$
A	altitude-high$\longrightarrow$	100.0	8954.7	1070.9
A	altitude-low$\longrightarrow$	42.0	9736.9	2198.8
C	continent-asia$\longrightarrow$	Asia	7019.4	1611.0
M	drymonths-few$\longrightarrow$	1.0	25083.8	7225.7
M	drymonths-many$\longrightarrow$	4.0	11524.2	1718.7
F	foresttype-dry$\longrightarrow$	Dry	11524.2	1718.7
F	foresttype-mangrove$\longrightarrow$	MoistMangrove	10632.2	6656.6
F	foresttype-moist$\longrightarrow$	Moist	3986.4	489.1
F	foresttype-wet$\longrightarrow$	Wet	18408.5	3090.1
O	oldgrowth-0$\longrightarrow$	0	11702.0	4628.4
R	rainfall-high$\longrightarrow$	3125.0	16162.3	2824.7
R	rainfall-low$\longrightarrow$	1862.0	11816.2	3965.9
-	random-1$\longrightarrow$	-	2400.3	126.8
-	random-2$\longrightarrow$	-	3954.3	1488.6

2018)). If it performs better than random it means that such a shift exists and is significant.

2.3.4 Experimental Evaluation of the Additional Error Model

Train and test datasets are created by randomly splitting the 1481 trees of Chave dataset into 741 (train) and 740 (test) trees. To create the necessary dataset to train the $\Delta_{s,t}MAE$ predictor, we create a synthetic training dataset by repeating the following operations 1000 times for random shifts and 2000 times for the 12 other shift types:

/ sum-
e syn-
dataset
on

- randomizing the trees from the train dataset
- creating source and target datasets based on the shift definition
- training a classical Chave model on the source dataset
- computing MAE and $\mathcal{H}Disc$ for source and target datasets
- the two previous steps are repeated by switching source and target datasets to create the reverse shifts ($\longleftarrow$)

In total, the training dataset is composed of 50,000 trees: 1000 trees for the random splits ($\longrightarrow$), 1000 trees for the reverse random splits ($\longleftarrow$), 24,000 trees for the 12 other splits ($\longrightarrow$) and 24,000 for their reversed ($\longleftarrow$). All models are trained on this dataset (Gradient Boosted Decision Trees Classifier, Ridge Regression for $\Delta_{s,t}MAE$ prediction).

For the evaluation, the test dataset is used. It is first divided into source and target datasets based on the shift definition. A Chave model is trained on the source dataset and MAE and $\Delta_{s,t}MAE$ are computed. These values are then compared to the ones predicted using the Ridge Regression predictor previously trained.

3 Results

3.1 Model Performance

As can be seen on Table 3, our proposed approach *COFARM-NN* reduces the MAE by 17% (or 31.5 kg per tree) compared to the original AM. It also reduces the RMSE and improves R^2 yet the result is not significant, possibly due to data size and inherent variance of the metric.

The *COFARM* model also improves significantly the test MAE by 14%, which indicates that the deep part of the *COFARM-NN* model accounts for an additional 3% reduction and confirms the usefulness of additional contextual information. Including this information into *ContextualChave* improved a bit the test MAE, however its performance stayed behind *COFARM-NN* ones.

LogReg performs comparably to the original AM which is expected as the log-log regression is equivalent to learning the power model from Chave et al. (2014). However adding deep layers in *LogReg-NN* doesn't produce gains vs *LogReg*, indicating that just increasing representation power is not sufficient to improve performance.

HGBRT, a different kind of method but having access to the same contextual information behaves surprisingly badly (just on par with a simple log-log linear regression), even though the model is considered a strong baseline in many regression problems (Shi et al., 2022). While the root cause is unclear, we believe it could be linked to inefficient regularization that leads to over-fitting, as indicated by a large difference between train and test metrics.

All in all, these results provide evidence of the generalization power of the *COFARM* and *COFARM-NN* models. Besides, the origin of the accuracy gains of *COFARM-NN* comes from better, more conservative calibration of predictions as can be observed on Figure 4. Indeed, the *COFARM-NN* calibration curve is closer to the ideal value of 1.0, especially so for small but also large AGBs (> exp(6) = 400 kgs), an area where the original AM is be suboptimal.

3.2 Evidence of Covariate Shift

Table 2 shows the Wasserstein distance for each shift type that has been implemented. The distance is the lowest for random shifts, meaning that there is little difference between source and target distributions for those shifts. $foresttype-moist$ shift shows also a low distance, meaning there is only a minor shift of distribution. $drymonths-few$ has the highest distance, highlighting a very different distribution for the trees located in places where there is no dry months. $\mathcal{H}Disc$ is also the highest for $drymonths-few$ showing the maximal discrepancy for this shift. However, there is no linear relationship between the Wasserstein distance and $\mathcal{H}Disc$, i.e. a higher distance is not always related to a higher $\mathcal{H}Disc$. This is probably because the distance is model-agnostic and

Table 3. Predictive power of different AMs (**bold** = best average, * = statistical significance at 90% vs Chave et al. (2014))

Dataset	Metric	(Chave et al., 2014)	*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	0.883 ± 0.029	**0.881** ± 0.026

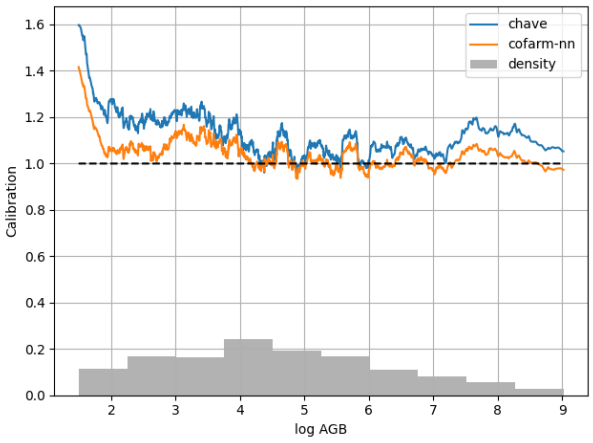

Figure 4. Calibration of AGB predictions

rely only on covariates distributions while $\mathcal{H}Disc$ is computed based on Chave model.

For all shifts, except the random ones, Gradient Boosting Classifiers are able to clearly predict that data from target datasets are different from the source ones (data not shown). This highlights a clear shift in the covariates distributions between the source and target distributions even when the Wasserstein distance is low. One can then argue about whether a model trained on the trees from a source dataset is applicable to another one.

3.3 Generalization Error Prediction

Figure 5 shows the predicted additional MAE due to distributions shift as a function of the observed one. Our model was able to predict $\Delta_{s,t}MAE$ with a high precision (R^2 = 0.832) meaning that it was able to learn the additional error based only on $\mathcal{H}Disc$ and $sign(\Delta_{s,t}Z)$. The red area represents the predictions that would have been very problematic because a positive $\Delta_{s,t}MAE$ was predicted as negative. However, this figure shows that our model is very conservative and is able to predict the sign and the magnitude of the additional error correctly. More details about MAE values for each shift can be found in Table A1.

The final model equation stands as follows:

$$\begin{aligned}\Delta_{s,t}MAE(h) = -19.69 \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h)\\ +13.72 \times sign(\Delta_{s,t}Z) + 44.86 \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h)^2\\ +153.80 \times (sign(\Delta_{s,t}Z) \times \mathcal{H}Disc_{\mathcal{H},L}(p_t, p_s, h))\end{aligned} \tag{8}$$

The coefficient for $\mathcal{H}Disc$ is negative, however the one for $\mathcal{H}Disc^2$ is twice as big and positive leading to a global positive effect of $\mathcal{H}Disc$ on $\Delta_{s,t}MAE$. This means that $\Delta_{s,t}MAE$ will get bigger when $\mathcal{H}Disc$ increases.

The coefficient for $sign(\Delta_{s,t}Z)$ being positive, it means that $\Delta_{s,t}MAE$ will be higher when $\mathbb{E}_{X\sim p_s}[D^2HW] > \mathbb{E}_{X\sim p_t}[D^2HW]$, i.e. when source covariates are bigger than target covariates.

The interaction term between $sign(\Delta_{s,t}Z)$ and $\mathcal{H}Disc$ shows the biggest coefficient highlighting that when $\mathbb{E}_{X\sim p_s}[D^2HW] > \mathbb{E}_{X\sim p_t}[D^2HW]$ a high $\mathcal{H}Disc$ will increase the $\Delta_{s,t}MAE$. On the contrary, when $\mathbb{E}_{X\sim p_s}[D^2HW] < \mathbb{E}_{X\sim p_t}[D^2HW]$ a high $\mathcal{H}Disc$ will decrease the $\Delta_{s,t}MAE$.

4 Discussion

4.1 Summary of key findings

We present a novel deep-learning model to predict AGB from easy to obtain tree measurements and contextual information. This model proves to reduce significantly predictions errors, especially for large trees. The proposed model is also able to incorporate additional information related to growth conditions. We also note that reduced error in AGB estimation should decrease uncertainty when estimated using Monte-Carlo noise injection methods as in (Réjou-Méchain et al., 2017).

Moreover, we propose a principled method to predict the additional error incurred by changing growth conditions between the site where a given AM is learned and the one where it is applied. This method can be used with any AM given that one has access to tree measurement data from both sites (AGB data is not required), allowing practitioners to test applicability and gauge from the data.

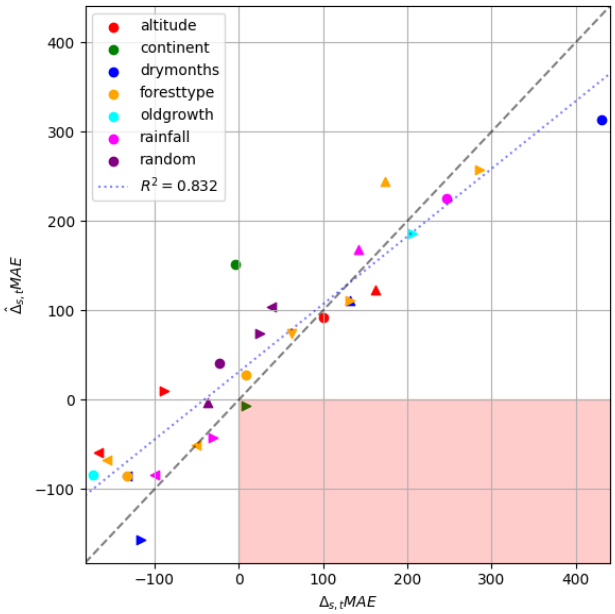

Figure 5. Predicted vs Actual Additional Error due to Covariate Shift. Color indicates the variable chosen to simulate the shift while marker shapes indicate different shifts magnitudes and directions.

4.2 Limitations

Limitations of present work include data, modeling and methodological considerations. First, we'd like to praise the authors of Chave et al. (2014) for the constitution of this incredibly helpful dataset. Yet, the datasets we use are still imperfect and may potentially misrepresent characteristics of tropical tree growth as they can be found in the wide area of tropical forests. In that respect we inherit the limitations of the original work, including the applicable range of D (5-156cm) and the exclusion of palms and liana species. More diverse open source biomass datasets would be required to develop more general pan-tropical AMs. Experiments on adding contextual information to AMs could be conducted only on the portion of the dataset that contains such information, totaling 1,481 trees. Second, even though we took care to evaluate a significant number of ML models, including linear, tree-based and deep neural networks, we possibly missed more accurate models. Finally, our proposed error prediction method is based on established statistical learning theory, yet the generalization bounds tightness can be questioned. Even if our experiments suggest a strong link and predictability in practice, only a larger scale use of the technique would give necessary feedback. Additional studies on other potential covariate shifts between conditions for learning and applying an AM would be of great value to the community.

4.3 Recommendations and Perspectives

4.3.1 Applicability of Allometric Models

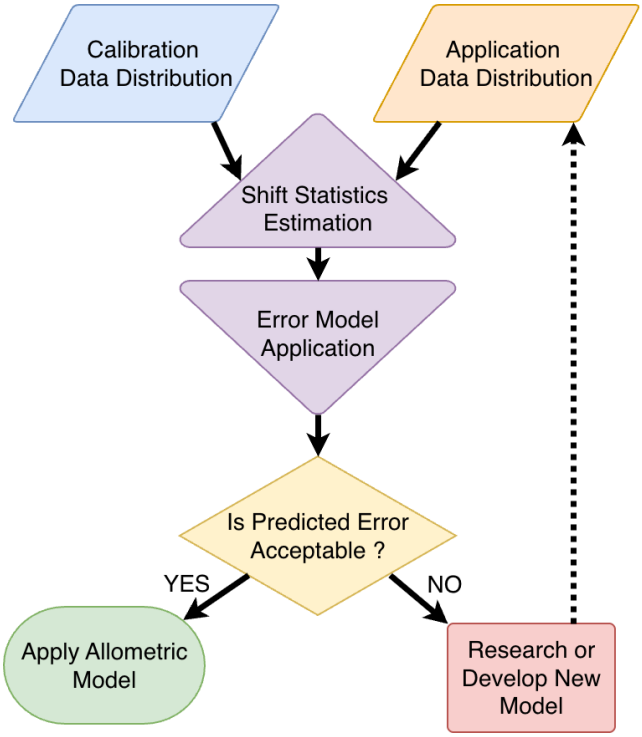

Figure 6. Recommended Process for Qualifying or Adjusting Allometric Models based on Predicted Additional Error.

In light of the experiments in Section 3.3, we propose a data-driven sanity check to decide if a given AM is suitable to be applied to a new dataset of tree measurements. The process is depicted if Figure 6 and starts with the computation of statistics on both the calibration dataset used to fit the AM and the newly considered dataset. It follows by injection of these statistics in the model of Equation 8 an estimation of the error incurred by the use of this AM in the new application site. The principal can then declare if the forecasted error falls within her pre-defined acceptable error range. If the answer is positive, the biomass predictions of the AM are judged satisfying and can be used. Conversely, an alternative AM should be considered; and if unfortunately none is available a case made to develop a new AM using destructive measurements. The advantage of this process are three-fold. First, it bases AM selection on objective, data-driven criteria. Second, it provides trust and visibility as shift statistics may be openly communicated to stakeholders. Finally, it motivates creation of new AM and the incurred costs on a strictly needed basis. We note that the proposed sanity check should be understood as a necessary yet not sufficient condition and by no means qualifies a given AM by itself, but rather adds an objective criteria of applicability. Indeed, the absence of

a distribution shift on tree measurements doesn't guarantee that the AM captured the relation of the former with AGB and we still recommend in addition to follow the prescriptions in each AM reference publication.

4.3.2 Future works

We foresee exciting future works in many directions. On the allometric modeling side, we see recent developments in Geospatial Foundation Models (Brown et al., 2025) as an opportunity to bring even more contextual information to AMs. We also anticipate that deep embeddings used to represent contextual information could well incorporate species information when it is available. On the methodological side, we encourage researchers in the field to get more into the habit of releasing primary data pertaining to AM development. Not only would it allow more modeling research to happen, but also we could increase reliability of AM choices by generalizing the kind of safety checks we propose to all AMs. More importantly, as a scientific field, open data is a formidable opportunity to develop better, unified models that one day could cover the whole spectrum of tropical forests, both primary and managed.

Author contributions. Conceptualization: E.D.; Methodology: E.D.; Software: A.D., E.D.; Experiments: A.D., E.D.; Formal analysis: E.D.; Writing: A.D., E.D.

Competing interests. A.D. and E.D are employed by PUR, a leading project implementer of Nature Based Solutions in which allometric equations may be used in the process of generating carbon credits under different certification standards. Note that the proposed model produces more conservative biomass estimates compared to established works.

Acknowledgements. Illustrations in Figure 2 composed using free to use images by *macrovector* on Freepik.

References

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Vaughan, J. W.: A theory of learning from different domains, Machine Learning, 79, 151–175, https://doi.org/10.1007/s10994-009-5152-4, 2010.

Blaufelder, C., Levy, C., Mannion, P., and Pinner, D.: A blueprint for scaling voluntary carbon markets to meet the climate challenge.

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., and Kohli, P.: AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data, https://doi.org/10.48550/arXiv.2507.22291, arXiv:2507.22291 [cs], 2025.

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.-P., Nelson, B. W., Ogawa, H., Puig, H., Riéra, B., and Yamakura, T.: Tree allometry and improved estimation of carbon stocks and balance in tropical forests, Oecologia, 145, 87–99, https://doi.org/10.1007/s00442-005-0100-x, 2005.

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., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C. M., Saldarriaga, J. G., and Vieilledent, G.: Improved allometric models to estimate the aboveground biomass of tropical trees, Global Change Biology, 20, 3177–3190, https://doi.org/10.1111/gcb.12629, 2014.

Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., and Shah, H.: Wide & Deep Learning for Recommender Systems, https://doi.org/10.48550/arXiv.1606.07792, arXiv:1606.07792 [cs], 2016.

Cortes, C. and Mohri, M.: Domain adaptation and sample bias correction theory and algorithm for regression, Theoretical Computer Science, 519, 103–126, https://doi.org/10.1016/j.tcs.2013.09.027, 2014.

Dutta Roy, A. and Debbarma, S.: Comparing the allometric model to machine learning algorithms for aboveground biomass estimation in tropical forests, Ecological Frontiers, 44, 1069–1078, https://doi.org/10.1016/j.ecofro.2024.05.010, 2024.

Friedman, J. H.: Stochastic gradient boosting, Computational Statistics & Data Analysis, 38, 367–378, https://doi.org/10.1016/S0167-9473(01)00065-2, 2002.

Gulrajani, I. and Lopez-Paz, D.: In Search of Lost Domain Generalization, https://doi.org/10.48550/arXiv.2007.01434, arXiv:2007.01434 [cs], 2020.

Henry, M., Bombelli, A., Trotta, C., Alessandrini, A., Birigazzi, L., Sola, G., Vieilledent, G., Santenoise, P., Longuetaud, F., Valentini, R., and others: GlobAllomeTree: international platform for tree allometric equations to support volume, biomass and carbon assessment, Iforest-biogeosciences and forestry, 6, 326, publisher: SISEF-Italian Society of Silviculture and Forest Ecology, 2013.

Hunka, N., Duncanson, L., Armston, J., Dubayah, R., Healey, S. P., Santoro, M., May, P., Araza, A., Bourgoin, C., Montesano, P. M., Neigh, C. S. R., Grantham, H., Potapov, P., Turubanova, S., Tyukavina, A., Richter, J., Harris, N., Urbazaev, M., Pascual, A., Suarez, D. R., Herold, M., Poulter, B., Wilson, S. N., Grassi, G., Federici, S., Sanz, M. J., and Melo, J.: Intergovernmental Panel on Climate Change (IPCC) Tier 1 forest biomass estimates from Earth Observation, Scientific Data, 11, 1127, https://doi.org/10.1038/s41597-024-03930-9, publisher: Nature Publishing Group, 2024.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, https://doi.org/10.48550/arXiv.1412.6980, arXiv:1412.6980 [cs], 2017.

Lopez-Paz, D. and Oquab, M.: Revisiting Classifier Two-Sample Tests, https://doi.org/10.48550/arXiv.1610.06545, arXiv:1610.06545 [stat], 2018.

Lovell, H. and MacKenzie, D.: Allometric equations and timber markets: An important forerunner of REDD+?, in: The Politics of Carbon Markets, Routledge, num Pages: 22, 2014.

Luxburg, U. v. and Schoelkopf, B.: Statistical Learning Theory: Models, Concepts, and Results, https://doi.org/10.48550/arXiv.0810.4752, arXiv:0810.4752 [stat], 2008.

Nocedal, J. and Wright, S. J.: Numerical optimization, Springer series in operations research, Springer, New York, 2nd ed edn., 2006.

Návar, J.: Measurement and assessment methods of forest aboveground biomass: A literature review and the challenges ahead, Biomass, pp. 27–64, publisher: Sciyo: Rijeka, Croatia, 2010.

Panaretos, V. M. and Zemel, Y.: Statistical Aspects of Wasserstein Distances, Annual Review of Statistics and Its Application, 6, 405–431, https://doi.org/10.1146/annurev-statistics-030718-104938, publisher: Annual Reviews, 2019.

Pargent, F., Pfisterer, F., Thomas, J., and Bischl, B.: Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features, Computational Statistics, 37, 2671–2692, https://doi.org/10.1007/s00180-022-01207-6, 2022.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, https://doi.org/10.48550/arXiv.1912.01703, arXiv:1912.01703 [cs], 2019.

Richard, G., de Mathelin, A., Hébrail, G., Mougeot, M., and Vayatis, N.: Unsupervised Multi-Source Domain Adaptation for Regression, in: Lecture Notes in Computer Science book series (LNCS, volume 12457), vol. 12457 of *Lecture Notes in Computer Science book series*, pp. 395–411, Springer International Publishing, Ghent, Belgium, https://doi.org/10.1007/978-3-030-67658-2_23, 2020.

Rojas-García, F., De Jong, B. H. J., Martínez-Zurimendí, P., and Paz-Pellat, F.: Database of 478 allometric equations to estimate biomass for Mexican trees and forests, Annals of Forest Science, 72, 835–864, https://doi.org/10.1007/s13595-015-0456-y, number: 6 Publisher: BioMed Central, 2015.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., and Hérault, B.: : an r package for estimating above-ground biomass and its uncertainty in tropical forests, Methods in Ecology and Evolution, 8, 1163–1167, https://doi.org/10.1111/2041-210X.12753, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12753, 2017.

Santoro, M. and Cartus, O.: ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2007, 2010, 2015, 2016, 2017, 2018, 2019, 2020, 2021 and 2022, v6.0, https://doi.org/10.5285/95913FFB6467447CA72C4E9D8CF30501, artwork Size: 11441 Files, 698903278165 B Pages: 11441 Files, 698903278165 B, 2025.

Shi, Y., Ke, G., Chen, Z., Zheng, S., and Liu, T.-Y.: Quantized Training of Gradient Boosting Decision Trees, in: Advances in Neural Information Processing Systems, vol. 35, pp. 18 822–18 833, Curran Associates, Inc., https://papers.nips.cc/paper_files/paper/2022/hash/77911ed9e6e864ca1a3d165b2c3cb258-Abstract.html, 2022.

Sileshi, G. W.: A critical review of forest biomass estimation models, common mistakes and corrective measures, Forest Ecology and Management, 329, 237–254, https://doi.org/10.1016/j.foreco.2014.06.026, 2014.

Smith, W. B. and Brand, G. J.: Allometric Biomass Equations for 98 Species of Herbs, Shrubs, and Small Trees, Research Note NC-299. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station, 299, https://doi.org/10.2737/NC-RN-299, 1983.

Van Breugel, M., Ransijn, J., Craven, D., Bongers, F., and Hall, J. S.: Estimating carbon stock in secondary forests: Decisions and uncertainties associated with allometric biomass models, Forest Ecology and Management, 262, 1648–1657, https://doi.org/10.1016/j.foreco.2011.07.018, 2011.

Wilson, G. and Cook, D. J.: A Survey of Unsupervised Deep Domain Adaptation, ACM Transactions on Intelligent Systems and Technology, 11, 1–46, https://doi.org/10.1145/3400066, 2020.

Wongchai, W., Onsree, T., Sukkam, N., Promwungkwa, A., and Tippayawong, N.: Machine learning models for estimating above ground biomass of fast growing trees, Expert Systems with Applications, 199, 117 186, https://doi.org/10.1016/j.eswa.2022.117186, 2022.

Zhao, F., Guo, Q., and Kelly, M.: Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA, Agricultural and Forest Meteorology, 165, 64–72, https://doi.org/10.1016/j.agrformet.2012.05.019, 2012.

Appendix A: Additional Results

Table A1. Forecast of Additional Biomass Prediction Error due to Covariate Shift (test dataset)

Shift	$\mathrm{E}[AGB_s]$	$\mathrm{E}[AGB_t]$	ΔZ	$\mathcal{H}Disc$	η	MAE_s^c	MAE_t^c	$\Delta_{s,t}MAE$	$\hat{\Delta}_{s,t}MAE$
altitude-high⟵	661.9	857.6	+3598.9	913.7	447.9	171.7	271.6	+99.8	+92.2
altitude-high⟶	857.6	661.9	-3598.9	1070.9	447.9	296.1	208.1	-87.9	+10.0
altitude-low⟵	772.4	371.0	-8793.6	3414.6	276.9	220.8	53.8	-167.0	-59.3
altitude-low⟶	371.0	772.4	+8793.6	2198.8	276.9	53.0	215.4	+162.4	+122.6
continent-asia⟵	594.4	912.7	+6569.2	3369.5	359.8	197.3	193.7	-3.6	+152.1
continent-asia⟶	912.7	594.4	-6569.2	1611.0	359.8	165.0	174.1	+9.1	-6.6
drymonths-few⟵	621.0	1210.6	+21406.9	8809.3	638.8	167.7	598.3	+430.6	+312.8
drymonths-few⟶	1210.6	621.0	-21406.9	7225.7	638.8	291.1	175.4	-115.7	-156.7
drymonths-many⟵	760.7	257.7	-11134.2	4350.1	273.7	200.7	67.5	-133.2	-85.0
drymonths-many⟶	257.7	760.7	+11134.2	1718.7	273.7	67.8	199.1	+131.3	+111.0
foresttype-dry⟵	760.7	257.7	-11134.2	4350.1	273.7	200.7	67.5	-133.2	-85.0
foresttype-dry⟶	257.7	760.7	+11134.2	1718.7	273.7	67.8	199.1	+131.3	+111.0
foresttype-mangrove⟵	704.5	222.7	-10262.3	3695.8	226.9	191.6	35.6	-156.0	-67.1
foresttype-mangrove⟶	222.7	704.5	+10262.3	6656.6	226.9	33.6	207.4	+173.9	+244.6
foresttype-moist⟵	573.3	737.8	+313.1	92.6	363.2	136.9	199.6	+62.7	+73.9
foresttype-moist⟶	737.8	573.3	-313.1	489.1	363.2	207.6	216.3	+8.7	+28.3
foresttype-wet⟵	615.5	1064.0	+16042.0	7068.6	506.7	172.5	458.0	+285.5	+257.2
foresttype-wet⟶	1064.0	615.5	-16042.0	3090.1	506.7	235.8	185.4	-50.4	-50.1
oldgrowth-0⟵	763.5	234.5	-11685.8	4332.3	247.2	208.3	35.9	-172.4	-84.5
oldgrowth-0⟶	234.5	763.5	+11685.8	4628.4	247.2	43.2	250.0	+206.8	+185.9
rainfall-high⟵	621.5	986.1	+13746.6	6027.8	477.7	175.6	421.6	+246.0	+225.8
rainfall-high⟶	986.1	621.5	-13746.6	2824.7	477.7	217.1	186.7	-30.5	-42.5
rainfall-low⟵	788.9	322.2	-11514.4	4312.0	308.9	199.9	100.5	-99.3	-83.9
rainfall-low⟶	322.2	788.9	+11514.4	3965.9	308.9	101.0	242.9	+141.9	+167.9
random-1⟵	538.3	588.3	-99.3	106.8	341.6	183.0	159.5	-23.5	+40.6
random-1⟶	588.3	538.3	+99.3	126.8	341.6	158.3	183.0	+24.7	+74.6
random-2⟵	676.3	754.5	+3138.5	1399.9	386.3	175.8	214.0	+38.2	+103.4
random-2⟶	754.5	676.3	-3138.5	1488.6	386.3	210.4	173.8	-36.6	-2.8