https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Attribution of aDGVM mismatches in biomass and tree cover across Africa

Simon Scheiter[1] and Liam Langan[1]

[1]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

Correspondence: Simon Scheiter (simon.scheiter@senckenberg.de)

Abstract. Dynamic global vegetation models (DGVMs) are commonly evaluated against satellite-derived products such as biomass or tree-cover. Yet, most studies provide error metrics and maps that reveal where models fail rather than why. Here, we develop a residual-analysis framework to diagnose systematic mismatches between aDGVM results and remotely sensed aboveground biomass and tree cover and attribute them to climate, fire, and human influence. We computed grid-cell residuals between aDGVM simulations and remote-sensing estimates and modeled these residuals using generalized additive models (GAMs) with bioclimatic variables, fire activity, human footprint and population density, as well as a spatial smooth. Models were checked for concurvity and basis-dimension adequacy, and evaluated using spatial block cross-validation. For aboveground biomass, a GAM with twelve smooth terms explained 80.3% of residual deviance (adjusted R^2=0.798), while for tree cover, a GAM with similar structure explained 78.5% of residual deviance (adjusted R^2=0.78). Strikingly, bioclimatic variables dominated residuals in 82.9% of AGB grid cells, not the processes known to be absent from aDGVM, such as nitrogen cycling and land use. This suggests that re-calibrating existing climate responses would reduce model bias more than adding missing process modules. Fire and human pressure played a proportionally larger role for tree cover (42.3% of grid cells) than for biomass (17.3%), reflecting distinct ecological controls on net primary production versus canopy dynamics. A persistent spatial signal after accounting for all predictors (unique R^2 = 11.0% AGB, 9.6% TC) points to unmeasured processes including soil properties, wild megafauna, and land-use history. The framework transforms descriptive DGVM error maps into quantitative, spatially explicit process attribution directly translating model-data mismatches into prioritized development targets for next-generation vegetation models.

1 Introduction

Dynamic global vegetation models (DGVMs) are powerful tools to project terrestrial carbon stocks and fluxes, vegetation distribution, and land-atmosphere feedbacks under past and future climate and land-use change (Prentice et al., 2007). Their performance is increasingly assessed using satellite-derived products such as aboveground biomass, productivity or tree cover, which offer global coverage and consistent observational data (Dantas de Paula et al., 2020). Many such evaluations report spatial patterns of correlations, mean biases, and biome-level summaries between DGVM outputs and remote-sensing estimates, often accompanied by maps of data-model differences (e.g., Scheiter et al., 2020; Kelley et al., 2013; Forkel et al., 2019).

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

While such model evaluation efforts revealed important priorities for model development and improvement, several limitations persist. Previous studies highlight where models over- or underestimate biomass, tree cover, or other variables, but they do not necessarily identify which processes, whether included or ignored in the model, drive these mismatches. DGVM-remote sensing mismatches are rarely treated as explicit residuals, modeled in response to environmental or anthropogenic drivers such as gradients in climate, disturbance, and human pressure. This hampers formal attribution of error sources and quantification of their relative contributions and an explicit interpretation in terms of uncertain or missing DGVM processes, such as fire, degradation, or plantations. Further, spatial autocorrelation in both model outputs and observational products is often ignored, leading to overoptimistic estimates of model performance and inflated significance of predictor effects. Finally, ignoring correlated environmental predictors in DGVM-data comparisons, for instance in temperature- and precipitation-related bioclimatic variables, and ignoring a systematic treatment of multi-collinearity may complicate inference about individual drivers.

Data-model residuals and mismatches indicate model limitations, and can as such be informative, because DGVM biases contain information about structural model deficiencies. For example, systematic overestimation of biomass in frequently burned savannas may reflect a fire module that fails to capture observed fire regimes, while deviations of biomass in heavily utilized landscapes may indicate missing representations of logging, degradation, or forest plantations. Similarly, consistent residual patterns along hydroclimatic gradients can reveal mis-specified model processes characterizing drought or wet-season responses. Modeling data-model residuals against climate, disturbance, and human-pressure gradients can thus provide a powerful diagnostic tool, transforming descriptive error maps into quantitative attribution of model limitations to under-constrained or even lacking processes.

Here, we developed such a residual-analysis framework and applied it to residuals between aDGVM simulations of aboveground biomass and tree cover (Scheiter and Higgins, 2009; Martens et al., 2021), and remotely-sensed aboveground biomass and tree cover (Ross et al., 2021). We calculated grid-cell residuals between model outputs and satellite-based estimates and modeled these residuals using GAMs with smooths for bioclimatic variables, burned area as a proxy for fire activity, human footprint, human population density, and livestock density as human-pressure indicators. The models were fitted with restricted maximum likelihood and penalization, checked for basis adequacy and concurvity, and evaluated under spatial block cross-validation to obtain realistic out-of-sample performance. The aDGVM has been developed for tropical and sub-tropical ecosystems (Scheiter and Higgins, 2009), and our analysis focuses on Africa, using aDGVM simulation results presented by Martens et al. (2021). We compared biomass and tree-cover residuals, allowing assessment of whether the same drivers explain residuals of both state variables.

This work addressed three major knowledge gaps and key questions: (1) Structure of residuals: are aDGVM-remote sensing residuals in biomass and tree cover random noise or systematically structured along climatic, fire and human pressure gradients? (2) Relative importance of drivers: which drivers (climate, fire, or human pressure) explain residual variation, and how do their contributions differ among biomes and between biomass and tree cover? (3) Spatial biases: how much spatial structure remains after accounting for observable drivers, and what does this imply about missing or mis-represented processes?

Correspondingly, we tested three hypotheses: (H1) Structure of residuals: biomass and tree-cover residuals display strong, non-random spatial structure and consistent variation along climate, fire, and human-pressure gradients, indicating systematic

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

aDGVM deficiencies rather than random errors. (H2) Disturbance and human dominance: after controlling for broad climate gradients, fire activity and human pressure explain a substantial fraction of residual variation, especially in savannas, characterized by strong vegetation-fire interactions and land use, reflecting under-representation of disturbance and degradation processes in the aDGVM. (H3) Spatial biases: even after accounting for climate, fire, and human influence, significant residual spatial structure persists at regional to biome scales, pointing to additional missing or mis-specified processes, such as soil constraints, land use history, or region-specific plant species communities.

By combining residual modeling with GAMs, systematic variable selection, and spatial block cross-validation, this study moves beyond descriptive DGVM-remote sensing comparisons and provides a transparent, quantitative basis for prioritizing model improvements. Although the analysis was conducted for a specific DGVM (the aDGVM), the African continent and specific remote-sensing products, the framework is generic and can be applied to other models, variables, and regions.

2 Materials and Methods

2.1 aDGVM model results

We used aboveground biomass and tree-cover outputs from aDGVM simulations conducted by Martens et al. (2021). As we did not run novel aDGVM simulations for this study, we do not provide a detailed model description but refer previous work (Scheiter and Higgins, 2009; Scheiter et al., 2012). Model simulations were conducted using downscaled, daily climate forcing based on MPI ESM at 0.5° spatial resolution (for details see Martens et al., 2021). For the analyses, we extracted annual values for tree cover and aboveground biomass and aggregated the data for the period 2001 to 2020 by calculating mean values. As we focus on aboveground tree biomass and tree cover, we included only grid cells with aboveground tree biomass greater than 1t/ha to remove deserts without tree biomass and potential biases in the GAMs. The biome type of each grid cell was derived in a post-processing step where aDGVM output was classified into desert, C_4 grassland, C_3 grassland, C_4 savanna, C_3 savanna, woodland and forest using C_3/C_4 grass biomass, and cover of forest and savanna tree PFTs (for details see Martens et al., 2021).

2.2 Aboveground biomass and tree cover and residuals

To calculate data-model residuals, we used satellite-derived gridded products of aboveground biomass and tree cover (Ross et al., 2021). Both variables were provided at 1 km spatial resolution, and resampled to the 0.5° resolution of aDGVM simulations, using the mean value.

For each grid cell i, we computed absolute residuals r_i as

$$r_i^{AGB} = AGB_i^{aDGVM} - AGB_i^{RS} \qquad \text{and} \qquad r_i^{TC} = TC_i^{aDGVM} - TC_i^{RS}, \tag{1}$$

where AGB and TC are aboveground biomass and tree cover, respectively, and $aDGVM$ and RS indicate aDGVM model results from Martens et al. (2021) and remote sensing data from Ross et al. (2021), respectively. Diagnostic plots of the residuals (Fig. S1) support the use of Gaussian GAMs with identity link as the main analysis, with some caveats. Raw AGB residuals

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

show a moderate right skew (skewness = 0.89) with heavy tails (excess kurtosis = 3.84), while raw TC residuals are more symmetric (skewness = −0.51, excess kurtosis = 0.59). We retained the Gaussian family for consistency between the AGB and TC models.

2.3 Explanatory variables

We used variables describing bioclimate, fire, direct human impacts, biome type and spatial structure as predictor variables in the GAMs. All products were aggregated and resampled to the 0.5° spatial resolution of aDGVM simulations using the mean value. Variables, except bioclimatic variables, are mapped in Fig. S2.

Climate: For aDGVM simulations, we used daily data of, among others, precipitation and temperature (see previous section). Such daily time series cannot be included directly into GAMs. We therefore calculated the common 19 bioclimatic variables (Booth et al., 2014) from daily time series. We selected daily data for the period 2001 to 2020, and calculated monthly means of temperature and precipitation across this period. Bioclimatic variables were then calculated using the 'dismo' R package (Hijmans et al., 2023) from the monthly means (see also Scheiter et al., 2025).

Fire: We used monthly burned area fraction for the period 2019 to 2024 from Sentinel3, SYN v1.1 (Chuvieco et al., 2016). The available monthly data for the entire period were averaged to obtain one data layer for fire activity.

Human Footprint Index (HFI): We used HFI for the year 2020 (Venter et al., 2016a, b; Mu et al., 2022; Sanderson et al., 2022). It describes the impact of humans on terrestrial ecosystems, considering for example settlements, land use types, transport routes or human population density. We used HFI data created with the second generation method, HD3-2.

Human population density: We used the gridded population of the world, version 4 product (Center For International Earth Science Information Network-C. I. E. S. I. N.-Columbia University, 2017). As the data spans a large range of values and is skewed, log-transformed data were used for our analysis.

Nitrogen deposition: We used N deposition data created by Lamarque et al. (2013a, b) for oxidized nitrogen (NOY) and reduced nitrogen (NHX) as provided in the ISIMIP repository. We downloaded annual data for the period 2006 to 2099, then extracted and averaged N deposition for the period 2006 to 2020 to maximize temporal overlap with aDGVM results. Data for NOY and NHX were added to obtain a single data layer for N deposition.

Domestic livestock (LS): Data for livestock were derived from the FAO gridded livestock density dataset (GLW4, Gilbert et al., 2015, 2018). We selected cattle and goat as those are the predominant domestic animals in African savanna rangelands. Data for both animals were added and log-transformed for the analyses to account for skewed distributions of population data.

Biome type: In Martens et al. (2021), aDGVM results were classified into seven different biome types using grass biomass and tree cover in a post-processing step (see section aDGVM model results). We used this modeled biome type as predictor in the GAMs. We acknowledge that using aDGVM-derived biome classifications as a predictor for aDGVM residuals introduces a degree of circularity, as the biome type reflects the same model outputs that generate the residuals. However, we consider this more suitable than observation-based biome products, which utilize different classification schemes, different biome types and different spatial resolutions that may not be consistent with the aDGVM's functional types. Using the aDGVM biome ensures that the predictor is spatially and conceptually consistent with the model being evaluated.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Spatial structure: Longitude and latitude were included via a two-dimensional smooth to account for spatial effects not explained by and of the other predictors.

2.4 Modeling

Residuals were modeled using a sequence of GAMs with thin-plate regression splines as implemented in the 'mgcv' R package (Marra and Wood, 2011; Wood, 2011, 2017). The final model was obtained by univariate screening of bioclimatic variables: each of the 19 bioclimatic variables was individually added to a base model, and the resulting Akaike Information Criterion (AIC) reductions were used to rank their individual contributions. This screening was necessary to reduce the number of predictors in the model by identifying those that explain most of the deviation and to reduce concurvity. We further compared different models by inclusion or exclusion of predictors or predictor combinations.

For the univariate screening, we used

$$r_i^{var} = f_1(HFI_i) + f_2(Pop_i) + f_3(Fire_i) + f_4(Ndep_i) + f_5(LS_i) + f_6(x_i, y_i) + biome_i + \epsilon_i, \quad (2)$$

as base model, where f_j are smooth functions, var is aboveground biomass or tree cover, and ϵ_i are Gaussian residuals. This base model includes biome type, fire and human impacts, and a spatial smooth, but no bioclimatic variables. Then, we included each of the 19 bioclimatic variables individually and tracked changes in the AIC. Continuous predictors were standardized, and diagnostics of pairwise correlations and concurvity were used to avoid redundant variables.

The largest AIC reductions for aboveground biomass arose from bio18 (Precipitation of Warmest Quarter, ΔAIC=-385.88, Table **??**), bio17 (Precipitation of Driest Quarter, ΔAIC=-338.68, and bio7 (Temperature Annual Range, ΔAIC=-294.67). For tree cover, bio12 (Annual Precipitation, ΔAIC=-557.57), bio6 (Minimum Temperature of Coldest Month, ΔAIC=-230.86) and bio5 (Max Temperature of Warmest Month, ΔAIC=-216.03) showed the strongest AIC reductions.

To account for ecological interpretability and relevance, and to use a common and consistent set of bioclimatic variables for both aboveground biomass and tree cover residuals, we selected six bioclimatic variables for the final GAM: bio1 (Annual Mean Temperature), bio4 (Temperature Seasonality), bio6 (Minimum Temperature of Coldest Month), bio12 (Annual Precipitation), bio15 (Precipitation Seasonality) and bio18 (Precipitation of Warmest Quarter). We concede that this selection was not fully objective and based on the univariate screening (Table S1). For example, AIC reductions of bio1 was not amongst the variables with the strongest reductions, yet, it is a very common variable in ecology and biogeography. Further, temperature seasonality (bio4) showed strong reductions in AIC for both biomass and tree cover. For consistency with temperature-related variables, particularly temperature seasonality (bio4), we included precipitation seasonality (bio15) despite moderate reductions in AIC. We also ensured that the model is balanced and contains three variables related to precipitation and temperature, respectively.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Hence, the final model was

$$r_i^{var} = f_1(HFI_i) + f_2(Pop_i) + f_3(Fire_i) + f_4(Ndep_i) + f_5(LS_i) + f_7(bio1_i) + f_8(bio4_i) + f_9(bio6_i) + f_{10}(bio12_i) + f_{11}(bio15_i) + f_{12}(bio18_i) + f_6(x_i, y_i) + biome_i + \epsilon_i. \quad (3)$$

Each univariate continuous predictor uses a thin-plate spline with an initial basis dimension $k = 10$, except $Ndep$ with $k = 20$ and the spatial smooth that uses a two-dimensional thin-plate spline with $k = 50$ to capture regional patterns. Appropriate values for k were selected using Effective Degrees of Freedom (edf) values as implemented in the 'mgcv' R package. Models were fitted with restricted maximum likelihood and variable selection to allow shrinkage of uninformative terms and penalize wiggles in the smoothing splines. We checked multicollinearity of predictor variables and found that worst-case concurvity was below 0.99 for all predictors.

We assessed the model performance for 10 different combinations of predictors. Therefore, predictors were grouped into climate (C: bio1, bio4, bio6, bio12, bio15, bio18), disturbance (D: BurnedArea), human pressure (H: HFI, PopDens, Ndep, Animals), spatial structure (S: two-dimensional smooth $s(x, y)$), and biome type (B: biome). Then GLMs were fitted including or excluding different groups (Tables S2, S3). To compare performance different models, we calculated R^2 values, deviance explained and RMSE for each model. This systematic assessment allowed identifying important sets of predictors.

To quantify the independent and shared contributions of each predictor group to residual variance, we performed a variance partitioning analysis. The four groups climate (C), disturbance (D), human pressure (H), and spatial structure (S) were included or excluded, while biome (B) was included as a baseline term in all models. We fitted all $2^4 = 16$ possible subset GAMs (including the biome-only baseline) and extracted the adjusted R^2 for each model. The unique contribution of each group was defined as the marginal R^2 gained when that group was added last to the full model, i.e., $R^2_{\text{full}} - R^2_{\text{full without group}}$. The shared or joint fraction was computed as the total explainable R^2 (full model minus baseline) minus the sum of all unique contributions, and captures variance that is co-explained by multiple groups along common ecological gradients.

Using the final model (M9 in Table S2, eq. 3), we predicted residuals for Africa and compared them to aDGVM-based residuals. We created partial dependency plots to analyze the relation between residuals and predictors.

2.5 Spatial block cross-validation

To obtain performance estimates under spatial dependence, we used spatial block cross-validation implemented in the 'blockCV' package (Valavi et al., 2019). For the analysis, the study region was partitioned into spatial blocks of a specified size (100 km^2), and blocks were randomly assigned to five folds, balancing the number of observations per fold. For each fold, the final GAM (M9) was fitted to all data outside the held-out blocks and predictions were generated for the test blocks. Performance metrics included root mean squared error (RMSE), coefficient of determination (R^2) between observed and predicted residuals, and mean bias, all computed on the held-out folds.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

2.6 Driver importance

We identified the most important predictor in each grid cell. Therefore, we extracted the absolute partial effect of each predictor from GAM predictions, and identified the variable with the highest effect in each grid cell. We mapped the spatial patterns of variable importance. To analyze if most important drivers differed between biomes, we calculated for each driver the number of grid cells per biome where the driver was most important.

3 Results

3.1 Patterns of residuals

Spatial patterns of aboveground biomass and tree cover residuals differed across Africa (Figs. 1, 2). For AGB, the aDGVM underestimates biomass in central African rainforests (negative residuals) and overestimates (positive residuals) it to the east of those forests. Overestimation also occurs in parts of the Sahel and in southern African woodlands and savannas.

Spatial patterns of tree cover residuals differ markedly from AGB residuals. In the central African rainforest belt, tree cover residuals are close to zero because both the aDGVM and the observations approach 100% canopy cover. To the east and south of the rainforest, the aDGVM substantially overestimates tree cover (positive residuals), that is, the model simulates more closed canopy than is observed in the transition zone from forest to savanna. Across the broader savanna belt of West, East and southern Africa, residuals show a spatially heterogeneous mix of over- and underestimation. Overall, the drivers of tree cover residuals in open vegetation are more variable and spatially structured than for AGB. In northern Africa, tree cover residuals are generally small in absolute terms given the low overall vegetation cover in arid zones.

The spatial smooth was among the most influential terms in both models. For AGB, the spatial smooth reached effective degrees of freedom of 47.9 (F = 53.9, $p < 0.0001$), and for TC the corresponding values was 46.1 (F = 25.7, $p < 0.0001$). Large effective degrees of freedom indicate that a large fraction of residual variation is organized in broad spatial patterns not fully captured by the included climate, fire, and anthropogenic predictors.

The base model used for univariate screening of bioclimatic variables (including biome, fire, anthropogenic factors and a spatial smooth, but excluding bioclimatic variables) explained 72.5% of AGB residual deviance (R^2 = 0.722) and 72.4% of TC residual deviance (R^2 = 0.721). Deviance explained, R^2 and RMSE varied both between different GAMs and between AGB and TC (Table 1). For instance, deviance explained ranged between 65.7% and 80.3% for AGB and between 35.1% and 78.5% for TC. Deviance explained and R^2 were consistently higher for AGB than for TC. Ranges of values of AGB and TC in aDGVM simulations and data are different such that RMSE is not directly comparable. Including biome type as predictor consistently increased model performance (e.g., between M0 and M1). The best GAM was the full model M9, including spatial effects, biome, climate, fire and anthropogenic factors (80.3% explained deviation and 0.798 R^2 for AGB and 78.5% explained deviation and 0.780 R^2 for TC). The spatial block cross-validation supported the high performance of the full models, with average R^2 values of 0.775 for AGB and 0.758 for TC across all five folds. Compared to the in-sample R^2 of 0.798 (AGB) and

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

0.780 (TC), the spatial CV values are approximately 0.02 lower, indicating that the models generalise well to spatially held-out regions and that overfitting to local spatial patterns is modest.

Spatial patterns of residuals predicted with GAMs agreed well with residuals derived from aDGVM and remote sensing (Fig. 3). For AGB, deviance between aDGVM and GAM residuals were mostly located in central African rainforest areas, while they were widely scattered without a clear pattern for tree cover. After fitting the final GAM M9, both AGB and TC residuals were near-symmetric (skewness = -0.30 and -0.23, respectively, Fig. S1), indicating that the Gaussian family in GAMs captured the central tendency adequately for both variables. However, AGB GAM residuals retain substantially heavy tails (excess kurtosis = 5.44, Shapiro-Wilk W = 0.93). TC GAM residuals are close to Gaussian (excess kurtosis = 1.31, W = 0.99), making the Gaussian assumption well-justified for tree cover. The heavy tails in AGB GAM residuals affect primarily the confidence intervals and p-values of smooth terms, while estimated smooth shapes and deviance explained are more robust to this violation.

Partial effects from the GAMs showed pronounced, non-linear associations between climate, disturbance, and anthropogenic predictors and DGVM residuals for AGB and TC across Africa (Table S5). For AGB, temperature (bio1, bio4, bio6) and precipitation variables (bio12, bio15, bio18) all had significant effects, with strong partial effects of high precipitation (bio12, F = 7.4) and wet-season precipitation (bio18, F = 20.8), indicating underestimation of biomass in wetter regions (Fig 4). Mean temperature (bio1, F = 31.8) showed changes from negative to positive partial effects, while minimum temperature of the coldest month (bio6, F = 42.4) showed the opposite response. For TC, climate responses were generally of lower amplitude. Annual precipitation (bio12, F = 31.8) showed a marked positive partial effect at intermediate to high values, revealing systematic underestimation of tree cover in humid climates (Fig 5). Minimum temperature of the coldest month (bio6, F = 9.5) showed changes from positive to negative partial effects.

Disturbance and anthropogenic variables displayed mostly monotonic partial responses in both models (Figs. 4, 5). Burned area showed the strongest non-bioclimatic signal for AGB (F = 56.4) and the second strongest for TC (F = 23.4). Residuals became increasingly negative with increasing burned area, indicating consistent overestimation of woody biomass and cover in frequently burned savannas and woodlands. This pattern implies that the aDGVM fire module does not reduce biomass and tree cover sufficiently in regions with high fire activity. Human Footprint Index had a strong negative effect in both models (AGB: F = 16.3; TC: F = 18.4), with residuals declining steeply above intermediate HFI values, suggesting that the model overestimates vegetation in heavily human-modified landscapes. Livestock density also showed negative partial effects (AGB: F = 14.8; TC: F = 10.5), consistent with under-representation of large-scale herbivory in the aDGVM. In contrast, population density and N deposition showed positive partial effects in both models, meaning that aDGVM underestimates biomass and tree cover in densely populated areas and regions with high N deposition. In both models, the spatial smooth retained strong large-scale structure, with broad regions of positive and negative partial effects, indicating substantial residual spatial organization beyond the included drivers.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

3.2 Drivers of aDGVM-data residuals

Our analysis revealed only weak agreement between the most important drivers of AGB and TC residuals (Fig. 6). Agreement was mostly found in southern Africa where bio6 (minimum temperature of coldest month) was the dominant driver. Overall, bio6 was the main driver for AGB (in 60.8% of grid cells, Table 2), followed by bio1 (mean annual temperature, 21.6% of grid cells) and, in areas neighboring central African rainforests, fire (9.7%) and drivers related to human impact. In contrast, for tree cover, anthropogenic drivers were dominant in most areas north and south of the rainforests (HFI 21.6% of grid cells, burned area 12.1%). In the tropical rainforests, bio15 (precipitation seasonality) was identified as dominant driver explaining tree cover residuals. In total, climate-related variables explained AGB residuals in 82.9% of the grid cells but only in 57.7% for tree cover.

Dominant drivers differed between biomes (Tables 2, 3). For instance, tree cover residuals in C_4 savannas were mainly explained by bio6, followed by HFI. In forests, residuals were mainly explained by the precipitation-related variables bio12 and bio15.

Variance partitioning revealed that the majority of explainable R^2 is shared among predictor groups rather than by any single group (Fig. 7, Table S4). Shared variance accounted for 78.5% of total explainable R^2 for AGB and 78.3% for TC, reflecting the strong co-linearity of climate, human pressure, disturbance, and spatial structure along Africa's main ecological gradients. Among the unique contributions, the spatial smooth had the largest independent effect (AGB: 11.0%; TC: 9.6% of explainable R^2), confirming that unmeasured spatially structured processes contribute substantially even after all predictors are accounted for. Climate had the second-largest unique contribution (AGB: 5.5%; TC: 7.3%), followed by human pressure (AGB: 2.9%; TC: 3.0%) and disturbance (AGB: 2.1%; TC: 1.8%). When considered in isolation, climate and spatial structure each explained the majority of variance individually (C alone: 75.4%/83.2%; S alone: 89.4%/85.7% of explainable R^2 for AGB/TC), but most of this is co-explained with the other groups. The biome-only baseline R^2 was 0.011 for AGB and 0.298 for TC, indicating that tree cover varies more systematically with biome type than aboveground biomass does.

4 Discussion

4.1 Structured residuals and driver attribution

The full GAMs explained 80.3% of AGB and 78.5% of TC residual deviance, and spatial block cross-validation confirmed that models generalize well to held-out regions ($\Delta R^2 \approx 0.02$). This supports the first hypothesis: aDGVM residuals are not random noise but display strong, non-random spatial structure and systematic variation along environmental and anthropogenic gradients. These findings indicate structured model deficiencies rather than random error. The strong differences between biomes further illustrate this structure: biome type alone explains 29.8% of TC but only 1.1% of AGB residual variance (model M1), confirming that TC residuals are more strongly biome-structured than AGB residuals. This large difference can be explained by the biome classification scheme applied to aDGVM results: this scheme is based on tree cover thresholds and grass biomass, while tree biomass is not used. Therefore, biomes reflect different ranges of tree cover. The dominant-driver

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

analysis and variance partitioning provide complementary perspectives on this structure. The dominant-driver map identifies where each group of drivers has its largest absolute effect, and variance partitioning quantifies what fraction of variance is uniquely attributable to each group. These perspectives diverge: climate was identified as dominant in 82.9% of AGB grid cells, yet its unique R^2 contribution was only 5.5%, because ≈78% of explainable variance is shared among all predictor groups. This result is a direct consequence of Africa's steep ecological gradient from equatorial rainforest to arid savanna, along which climate, fire activity, and human pressure are all strongly correlated.

The second hypothesis is supported but the magnitude of fire and human impact is lower than expected. Fire and human pressure together dominate TC residuals in 42.3% of grid cells and AGB residuals in only 17.3%. Fire was the dominant driver in only 9.7% of AGB and 12.1% of TC grid cells. Two explanations are plausible. First, climate-driven model uncertainties may be large enough to dominate the residual signal, leaving limited variance for other factors. Second, the aDGVM fire module may already capture the most important patterns of fire impacts adequately. In either case, the finding reframes development priorities: since climate-calibration errors account for the majority of model bias, improving hydroclimate sensitivities would yield larger bias reductions than adding missing process modules such as nitrogen cycling or land use alone. The contrasting results for AGB and TC are ecologically meaningful. AGB is the direct result of net primary production simulated by aDGVM and is tightly coupled to temperature and water availability. Climate biases propagate directly into AGB residuals. Tree cover, in contrast, is derived from simulated crown areas and a heuristic factor describing density-dependent crown overlap, and is influenced by the ratio of savanna and forest trees due to their different crown architectures. This introduces model-specific uncertainty that may amplify the role of unresolved variation. Generally, tree cover emerges from a dynamic equilibrium between tree growth, fire-driven mortality, and herbivory that can sustain alternative stable states (Staver et al., 2011; Hirota et al., 2011; Higgins et al., 2023). Above ≈650 mm mean annual precipitation, disturbance regimes rather than rainfall become the primary regulators of tree cover (Sankaran et al., 2005), which explains why a larger share of TC residuals reflects fire and human pressure.

The third hypothesis is also confirmed. Even with all predictors included, the spatial smooth retains large effective degrees of freedom (AGB: 47.9; TC: 46.1), and its unique R^2 contribution (11.0% AGB, 9.6% TC) exceeds that of any other predictor group. This persistent spatial structure points to missing processes that are spatially organized but not captured by the included variables, including soil physical properties (Higgins et al., 2023), groundwater access, wild megafauna (Pachzelt et al., 2015), and continental-scale land-use history. Additionally, this spatial variance likely captures the regional-scale legacy effects of the alternative stable states (Moncrieff et al., 2014), where localized fire histories or past disturbances have locked vegetation into specific structural states independently of current macro-environmental drivers (e.g., Archibald et al., 2013; Case and Staver, 2017).

4.2 GAMs as diagnostic tools for DGVMs

We used GAMs to model residuals between aDGVM results and remote sensing data. In contrast to machine learning approaches such as random forests, GAMs produce directly interpretable smooth functions and partial dependency, enabling rigorous process attribution and realistic extrapolation (Wood, 2017; Pedersen et al., 2019). Treating data-model differences

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

as residuals to be modeled, rather than as descriptive error maps, transforms DGVM evaluation from ranking to diagnosis (Langan et al., 2025). Thereby, the combination of systematic variable selection, concurvity diagnostics, and spatial block cross-validation is essential. Random cross-validation would yield optimistic performance estimates under spatial autocorrelation, whereas spatial block CV provides realistic out-of-sample errors and guards against interpreting spatially autocorrelated noise as explained variance.

Our framework is generic and not limited to aDGVM or Africa. Applying the same approach to other models such as LPJ-GUESS or CLM would allow the community to distinguish model-specific from general process deficiencies. Our findings are broadly consistent with evaluations of other DGVMs, which report analogous patterns of overestimation of biomass in fire-prone savannas and underestimation in humid forests (Kelley et al., 2013; Forkel et al., 2019), suggesting that the identified deficiencies reflect structural limitations common to current-generation DGVMs.

4.3 Implications for aDGVM development

The partial effects and driver importance analyses identify five priority areas for aDGVM improvement. First, and most critically, hydroclimate calibration: bioclimatic variables dominate residuals across the continent. The model underestimates biomass and tree cover in humid regions (high annual precipitation, bio12, and high precipitation of warmest quarter bio18) and shows reversal of the sign of effects of mean annual temperature (bio1) and minimum temperature of the coldest month (bio6) responses, indicating that drought sensitivity and wet-season productivity are systematically mis-calibrated. Because climate errors account for the majority of the total bias, improving these sensitivities offers the largest potential reduction in model-observation mismatch. Second, fire: residuals decrease strongly with burned area, indicating that the fire module does not reduce biomass and tree cover sufficiently in high-fire-frequency regions. Third, nitrogen cycling: N deposition shows positive partial effects in both models, meaning the model underestimates productivity in high-N-deposition regions, a consequence of the absence of a nitrogen cycle in aDGVM. Fourth, land use and herbivory: negative effects of HFI and livestock density point to insufficient representation of land degradation and grazing. While land-use processes have been implemented in aDGVM at regional scale (Scheiter and Savadogo, 2016; Scheiter et al., 2019), continental-scale parameterization is lacking. Wild megafauna, such as elephants, buffalo, or wildebeest, additionally exert browsing and grazing impacts that exceed livestock density in many protected areas (Sankaran et al., 2008) and were only included at local scale in the aDGVM (Scheiter and Higgins, 2012). Wild animals were included in the LPJ-GUESS but their impact on biome patterns and carbon was low (Pachzelt et al., 2015). Fifth, soil properties and plant hydraulics: the remaining spatial signal likely reflects missing edaphic and hydraulic controls on woody cover, including soil texture and water-holding capacity. While soil properties are included in aDGVM simulations, we did not include them as predictors in our GAM analyses.

The importance of main drivers of aDGVM-data residuals varied systematically across biomes. TC residuals in C_4 savannas were primarily explained by bio6 and HFI, whereas in forests precipitation-related variables (bio12, bio15) dominated. This biome-specificity highlights that no single process improvement will uniformly reduce aDGVM bias across Africa; targeted interventions are needed for different ecosystem types.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

4.4 Caveats and limitations

Our approach has several limitations. The remote-sensing benchmark products carry spatially structured uncertainty. Signal saturation in radar-based retrievals can lead to underestimation of AGB in dense tropical forests, while spectral similarity between tree canopy, tall shrubs, and senescent grass introduces errors in semiarid systems. Because these uncertainties are spatially structured, they may generate patterns in the GAM residual surface that reflect measurement error rather than aDGVM process deficiencies. Future analyses would benefit from propagating product-specific uncertainty estimates directly into the residual model.

The GAM framework assumes Gaussian errors; however, the AGB model residuals show excess kurtosis (5.44; Shapiro–Wilk $W = 0.93$), indicating heavier-than-Gaussian tails. Confidence intervals and p-values for GAM smooth terms may therefore be slightly anti-conservative and should be interpreted with caution.

The bioclimatic predictor variables were derived from the same climate forcing used to drive aDGVM (MPI-ESM, Martens et al., 2021) rather than from observational products such as CHELSA (Karger et al., 2017) or WorldClim (Fick and Hijmans, 2017). Any systematic biases in climate data used as driver in aDGVM therefore affect both the residuals and the predictor values, potentially obscuring the true relationship between observed climate and model performance. Similarly, biomes were derived from aDGVM simulations and may therefore bias the contribution of biome in the GAMs.

5 Conclusions

This study demonstrates that aDGVM residuals in aboveground biomass and tree cover across Africa are not random noise but carry structured, interpretable signals of model deficiencies. By modeling these residuals with GAMs against climate, fire, human pressure, and spatial structure, we transform a standard model evaluation exercise into a quantitative diagnostic. Climate mismatches dominate AGB errors (82.9% of grid cells), while fire, human pressure, and herbivory play a proportionally larger role for tree cover (42.3% of grid cells), reflecting the distinct ecological controls on each variable. Notably, climate errors, not the processes known to be absent from aDGVM such as nitrogen cycling and land use, dominate the model bias. This implies that re-calibrating existing relations between vegetation and hydroclimate would reduce the mismatch more than adding missing process modules alone. Soil properties, wild megafauna, and continental-scale land use emerge as the additional missing processes responsible for the persistent spatial bias that remains after all predictors are accounted for. The framework is not limited to aDGVM or Africa; it is generic, computationally tractable, and directly translates model-data mismatches into prioritized development targets. As satellite-derived benchmark datasets continue to improve in spatial resolution, temporal depth, and uncertainty quantification, residual-based GAM frameworks of this kind offer a powerful pathway for turning the growing volume of Earth observation data into actionable guidance for next-generation vegetation model development, with direct implications for the accuracy of terrestrial carbon cycle projections and climate change assessments.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Code and data availability. (1) aDGVM results and bioclimatic variables are provided in https://doi.org/10.5281/zenodo.20695900, Scheiter and Langan (2026)

(2) Ross et al. tree cover and aboveground biomass: https://doi.org/10.3334/ORNLDAAC/1777, Ross et al. (2021)

(3) ESA burned area: https://data.ceda.ac.uk/neodc/esacci/fire/data/burned_area/Sentinel3_SYN/grid/v1.1/, Chuvieco et al. (2016)

(4) Human footprint index: https://wcshumanfootprint.org/data-access, Venter et al. (2016b); Mu et al. (2022); Sanderson et al. (2022)

(5) Human population density: https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-gpwv4-popdens-r11-4.11, Center For International Earth Science Information Network-C. I. E. S. I. N.-Columbia University (2017)

(6) Livestock density: https://data.apps.fao.org/catalog/iso/9d1e149b-d63f-4213-978b-317a8eb42d02, Gilbert et al. (2015, 2018)

(7) Nitrogen deposition: https://data.isimip.org, Lamarque et al. (2013a, b)

(8) R scripts: https://doi.org/10.5281/zenodo.20695900, Scheiter and Langan (2026)

Author contributions. SS conceived the study, conducted the analyses and lead the writing of the manuscript; LL reviewed and edited the manuscript

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. We thank Loris Schmidt for creating initial versions of the R scripts. R scripts, statistical analyses, and table editing were developed with assistance from Claude Code (Anthropic) and Perplexity AI; the underlying scientific logic and data interpretation were independently verified by the authors. Claude Code was additionally used for language editing and improving the readability of parts of the manuscript text. All content was critically reviewed, revised, and verified by the authors, who take full responsibility for the final text.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

References

Archibald, S., Lehmann, C. E. R., Gómez-Dans, J. L., and Bradstock, R. A.: Defining pyromes and global syndromes of fire regimes, Proceedings of the National Academy of Sciences, 110, 6442–6447, https://doi.org/10.1073/pnas.1211466110, 2013.

Booth, T. H., Nix, H. A., Busby, J. R., and Hutchinson, M. F.: bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies, Diversity and Distributions, 20, 1–9, https://doi.org/10.1111/ddi.12144, 2014.

Case, M. F. and Staver, A. C.: Fire prevents woody encroachment only at higher-than-historical frequencies in a South African savanna, Journal of Applied Ecology, 54, 955–962, https://doi.org/10.1111/1365-2664.12805, 2017.

Center For International Earth Science Information Network-C. I. E. S. I. N.-Columbia University: Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11 (Version 4.11), https://doi.org/10.7927/H49C6VHW, 2017.

Chuvieco, E., Pettinari, M. L., and E. S. A. Fire C. C. I. project team: ESA Fire Climate Change Initiative (Fire CCI) : Sentinel-3 SYN Burned Area Grid product, version 1.1, https://catalogue.ceda.ac.uk/uuid/bcef9e87740e4cbabc743d295afbe849/, 2016.

Dantas de Paula, M., Gomez Gimenez, M., Niamir, A., Thurner, M., and Hickler, T.: Combining European Earth Observation products with Dynamic Global Vegetation Models for estimating Essential Biodiversity Variables, International Journal of Digital Earth, 13, 262–277, https://doi.org/10.1080/17538947.2019.1597187, 2020.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1km spatial resolution climate surfaces for global land areas, International Journal of Climatology, 37, 4302–4315, 2017.

Forkel, M., Drüke, M., Thurner, M., Dorigo, W., Schaphoff, S., Thonicke, K., von Bloh, W., and Carvalhais, N.: Constraining modelled global vegetation dynamics and carbon turnover using multiple satellite observations, Scientific Reports, 9, 18 757, https://doi.org/10.1038/s41598-019-55187-7, 2019.

Gilbert, M., Conchedda, G., Van Boeckel, T. P., Cinardi, G., Linard, C., Nicolas, G., Thanapongtharm, W., D'Aietti, L., Wint, W., Newman, S. H., and Robinson, T. P.: Income Disparities and the Global Distribution of Intensively Farmed Chicken and Pigs, PLOS ONE, 10, e0133 381, https://doi.org/10.1371/journal.pone.0133381, 2015.

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G. R. W., and Robinson, T. P.: Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010, Scientific Data, 5, 180 227, https://doi.org/10.1038/sdata.2018.227, 2018.

Higgins, S. I., Conradi, T., Kruger, L. M., O'Hara, R. B., and Slingsby, J. A.: Limited climatic space for alternative ecosystem states in Africa, Science, 380, 1038–1042, https://doi.org/10.1126/science.add5190, 2023.

Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J.: dismo: Species Distribution Modeling, https://CRAN.R-project.org/package=dismo, r package version 1.3-14, 2023.

Hirota, M., Holmgren, M., Van Nes, E. H., and Scheffer, M.: Global Resilience of Tropical Forest and Savanna to Critical Transitions, Science, 334, 232–235, https://doi.org/10.1126/science.1210657, 2011.

Karger, D. N., Conrad, O., B"ohner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas, Scientific Data, 4, 170 122, https://doi.org/10.1038/sdata.2017.122, 2017.

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., and Willis, K. O.: A comprehensive benchmarking system for evaluating global vegetation models, Biogeosciences, 10, 3313–3340, https://doi.org/10.5194/bg-10-3313-2013, 2013.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Lamarque, J.-F., Dentener, F., McConnell, J., Ro, C.-U., Shaw, M., Vet, R., Bergmann, D., Cameron-Smith, P., Dalsoren, S., Doherty, R., Faluvegi, G., Ghan, S. J., Josse, B., Lee, Y. H., MacKenzie, I. A., Plummer, D., Shindell, D. T., Skeie, R. B., Stevenson, D. S., Strode, S., Zeng, G., Curran, M., Dahl-Jensen, D., Das, S., Fritzsche, D., and Nolan, M.: Multi-model mean nitrogen and sulfur deposition from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): evaluation of historical and projected future changes, Atmospheric Chemistry and Physics, 13, 7997–8018, https://doi.org/10.5194/acp-13-7997-2013, 2013a.

Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P. J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S. J., Horowitz, L. W., Lee, Y. H., MacKenzie, I. A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S. T., Schulz, M., Skeie, R. B., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., and Zeng, G.: The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations and climate diagnostics, Geoscientific Model Development, 6, 179–206, https://doi.org/10.5194/gmd-6-179-2013, 2013b.

Langan, L., Kattenborn, T., Römermann, C., Scheiter, S., Anders, T., Wolf, S., Dantas de Paula, M., and Hickler, T.: Prediction in trait-based ecology: global simulations of specific leaf area using a trait-based dynamic vegetation model, bioRxiv, p. 2025.04.14.648688, https://doi.org/10.1101/2025.04.14.648688, 2025.

Marra, G. and Wood, S. N.: Practical variable selection for generalized additive models, Computational Statistics & Data Analysis, 55, 2372–2387, https://doi.org/10.1016/j.csda.2011.02.004, 2011.

Martens, C., Hickler, T., Davis-Reddy, C., Engelbrecht, F., Higgins, S. I., von Maltitz, G. P., Midgley, G. F., Pfeiffer, M., and Scheiter, S.: Large uncertainties in future biome changes in Africa call for flexible climate adaptation strategies, Global Change Biology, 27, 340–358, https://doi.org/10.1111/gcb.15390, 2021.

Moncrieff, G. R., Scheiter, S., Bond, W. J., and Higgins, S. I.: Increasing atmospheric CO_2 overrides the historical legacy of multiple stable biome states in Africa, New Phytologist, 201, 908–915, https://doi.org/10.1111/nph.12551, 2014.

Mu, H., Li, X., Wen, Y., Huang, J., Du, Z., Ma, P., Li, G., Guo, M., Yin, S., Gang, C., He, B., Zhuang, Q., Wu, B., Liu, Q., Gong, P., and Bryan, B. A.: A global record of annual terrestrial Human Footprint dataset from 2000 to 2018, Scientific Data, 9, 176, https://doi.org/10.1038/s41597-022-01284-8, 2022.

Pachzelt, A., Forrest, M., Rammig, A., Higgins, S. I., and Hickler, T.: Potential impact of large ungulate grazers on African vegetation, carbon storage and fire regimes, Global Ecology and Biogeography, 24, 991–1002, https://doi.org/10.1111/geb.12313, 2015.

Pedersen, E. J., Miller, D. L., Simpson, G. L., and Ross, N.: Hierarchical generalized additive models in ecology: an introduction with mgcv, PeerJ, 7, e6876, https://doi.org/10.7717/peerj.6876, 2019.

Prentice, I. C., Bondeau, A., Cramer, W., Harrison, S. P., Hickler, T., Lucht, W., Sitch, S., Smith, B., and Sykes, M. T.: Terrestrial Ecosystems in a Changing World, chap. Dynamic Global Vegetation Modeling: Quantifying Terrestrial Ecosystem Responses to Large-Scale Environmental Change, pp. 175–192, Springer, editors: Canadell, J. G. and Pataki, D. E. and Pitelka, L. F., 2007.

Ross, C. W., Hanan, N. P., Prihodko, L., Anchang, J., Ji, W., and Yu, Q.: Woody-biomass projections and drivers of change in sub-Saharan Africa, Nature Climate Change, 11, 449–455, https://doi.org/10.1038/s41558-021-01034-5, 2021.

Sanderson, E. W., Fisher, K., Robinson, N., Sampson, D., Duncan, A., and Royte, L.: The March of the Human Footprint, EcoEvoRxiv Pre-print, https://doi.org/10.32942/osf.io/d7rh6, 2022.

Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., Gignoux, J., Higgins, S. I., Roux, X. L., Ludwig, F., Ardo, J., Banyikwa, F., Bronn, A., Bucini, G., Caylor, K. K., Coughenour, M. B., Diouf, A., Ekaya, W., Feral, C. J., February, E. C., Frost, P. G. H., Hiernaux, P., Hrabar, H., Metzger, K. L., Prins, H. H. T., Ringrose, S., Sea, W., Tews, J., Worden, J., and Zambatis, N.: Determinants of woody cover in African savannas., Nature, 438, 846–849, https://doi.org/10.1038/nature04070, 2005.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Sankaran, M., Ratnam, J., and Hanan, N.: Woody cover in African savannas: the role of resources, fire and herbivory, Global Ecology and Biogeography, 17, 236–245, 2008.

Scheiter, S. and Higgins, S. I.: Impacts of climate change on the vegetation of Africa: an adaptive dynamic vegetation modelling approach (aDGVM), Global Change Biology, 15, 2224–2246, 2009.

Scheiter, S. and Higgins, S. I.: How many elephants can you fit into a conservation area, Conservation Letters, 5, 176–185, https://doi.org/10.1111/j.1755-263X.2012.00225.x, 2012.

Scheiter, S. and Langan, L.: Attribution of aDGVM mismatches in biomass and tree cover across Africa, https://doi.org/10.5281/zenodo.20695900, 2026.

Scheiter, S. and Savadogo, P.: Ecosystem management can mitigate vegetation shifts induced by climate change in West Africa, Ecological Modelling, 332, 19–27, https://doi.org/10.1016/j.ecolmodel.2016.03.022, 2016.

Scheiter, S., Higgins, S. I., Osborne, C. P., Bradshaw, C., Lunt, D., Ripley, B. S., Taylor, L. L., and Beerling, D. J.: Fire and fire-adapted vegetation promoted C_4 expansion in the late Miocene, New Phytologist, 195, 653–666, https://doi.org/10.1111/j.1469-8137.2012.04202.x, 2012.

Scheiter, S., Schulte, J., Pfeiffer, M., Martens, C., Erasmus, B. F. N., and Twine, W. C.: How Does Climate Change Influence the Economic Value of Ecosystem Services in Savanna Rangelands?, Ecological Economics, 157, 342 – 356, https://doi.org/https://doi.org/10.1016/j.ecolecon.2018.11.015, 2019.

Scheiter, S., Kumar, D., Corlett, R. T., Gaillard, C., Langan, L., Lapuz, R. S., Martens, C., Pfeiffer, M., and Tomlinson, K. W.: Climate change promotes transitions to tall evergreen vegetation in tropical Asia, Global Change Biology, 26, 5106–5124, https://doi.org/10.1111/gcb.15217, 2020.

Scheiter, S., Merkel, H., and Hickler, T.: Divergent projections of future African biome shifts with process-based and species distribution models, Ecography, 2025, e08 176, https://doi.org/10.1002/ecog.08176, 2025.

Staver, A. C., Archibald, S., and Levin, S.: Tree cover in sub-Saharan Africa: Rainfall and fire constrain forest and savanna as alternative stable states, Ecology, 92, 1063–1072, 2011.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G.: blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models, Methods in Ecology and Evolution, 10, 225–232, https://doi.org/10.1111/2041-210X.13107, 2019.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., and Watson, J. E. M.: Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation, Nature Communications, 7, 12 558, https://doi.org/10.1038/ncomms12558, 2016a.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., and Watson, J. E. M.: Data from: Global terrestrial Human Footprint maps for 1993 and 2009, https://doi.org/10.5061/dryad.052q5, 2016b.

Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73, 3–36, https://doi.org/10.1111/j.1467-9868.2010.00749.x, 2011.

Wood, S. N.: Generalized Additive Models: An Introduction with R, Chapman and Hall/CRC, 2nd edn., https://www.taylorfrancis.com, 2017.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Table 1. Comparison of different GAMs for biomass and tree cover, see Table X for definition of models M0 to M9. 'DevExpl' is deviance explained by the GAM, 'RMSE' is root mean squared error.

Model	AGB			TC		
	DevExpl	R^2	RMSE	DevExpl	R^2	RMSE
M0	0.657	0.654	31.2	0.351	0.346	20.3
M1	0.718	0.715	28.3	0.714	0.711	13.5
M2	0.726	0.722	27.9	0.451	0.443	18.7
M3	0.771	0.767	25.5	0.766	0.762	12.2
M4	0.681	0.676	30.1	0.379	0.371	19.9
M5	0.739	0.735	27.2	0.737	0.733	12.9
M6	0.700	0.696	29.2	0.390	0.383	19.7
M7	0.758	0.754	26.2	0.749	0.745	12.6
M8	0.759	0.754	26.2	0.475	0.464	18.3
M9	0.803	0.798	23.7	0.785	0.780	11.7

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Table 2. Driver importance per biome. Percent of grid cells per biome where different drivers are most important to explain AGB residuals.

Biome	bio1	bio4	bio6	bio12	bio18	BA	HFI	LS	Ndep	PD
all	21.62	0.23	60.83	0.02	0.20	9.71	3.50	1.05	2.69	0.17
Desert	16.67	0.00	83.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C4 gr	33.33	0.00	62.50	0.00	0.00	3.65	0.52	0.00	0.00	0.00
C3 gr	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C4 sav	26.26	0.03	64.28	0.00	0.03	5.96	2.02	0.80	0.52	0.07
C3 sav	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Woodl	8.62	0.42	70.21	0.00	0.50	13.89	3.93	1.00	1.17	0.25
Forest	21.85	0.46	47.90	0.06	0.29	13.86	6.04	1.61	7.65	0.29

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

Table 3. Driver importance per biome. Percent of grid cells per biome where different drivers are most important to explain tree cover residuals.

Biome	bio6	bio12	bio15	bio18	BA	HFI	LS	Ndep	PD
all	21.47	20.09	14.65	1.51	12.11	21.62	1.36	1.73	5.46
Desert	66.67	0.00	16.67	0.00	0.00	16.67	0.00	0.00	0.00
C4 gr	52.08	0.00	31.25	0.00	6.25	4.17	1.56	0.00	4.69
C3 gr	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C4 sav	34.08	13.43	11.75	0.17	10.95	23.58	2.20	0.59	3.24
C3 sav	40.00	0.00	0.00	0.00	0.00	30.00	0.00	0.00	30.00
Woodl	12.80	27.36	7.11	2.01	15.98	22.85	0.59	3.43	7.87
Forest	2.42	28.69	22.89	3.57	12.25	19.55	0.52	2.65	7.48

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

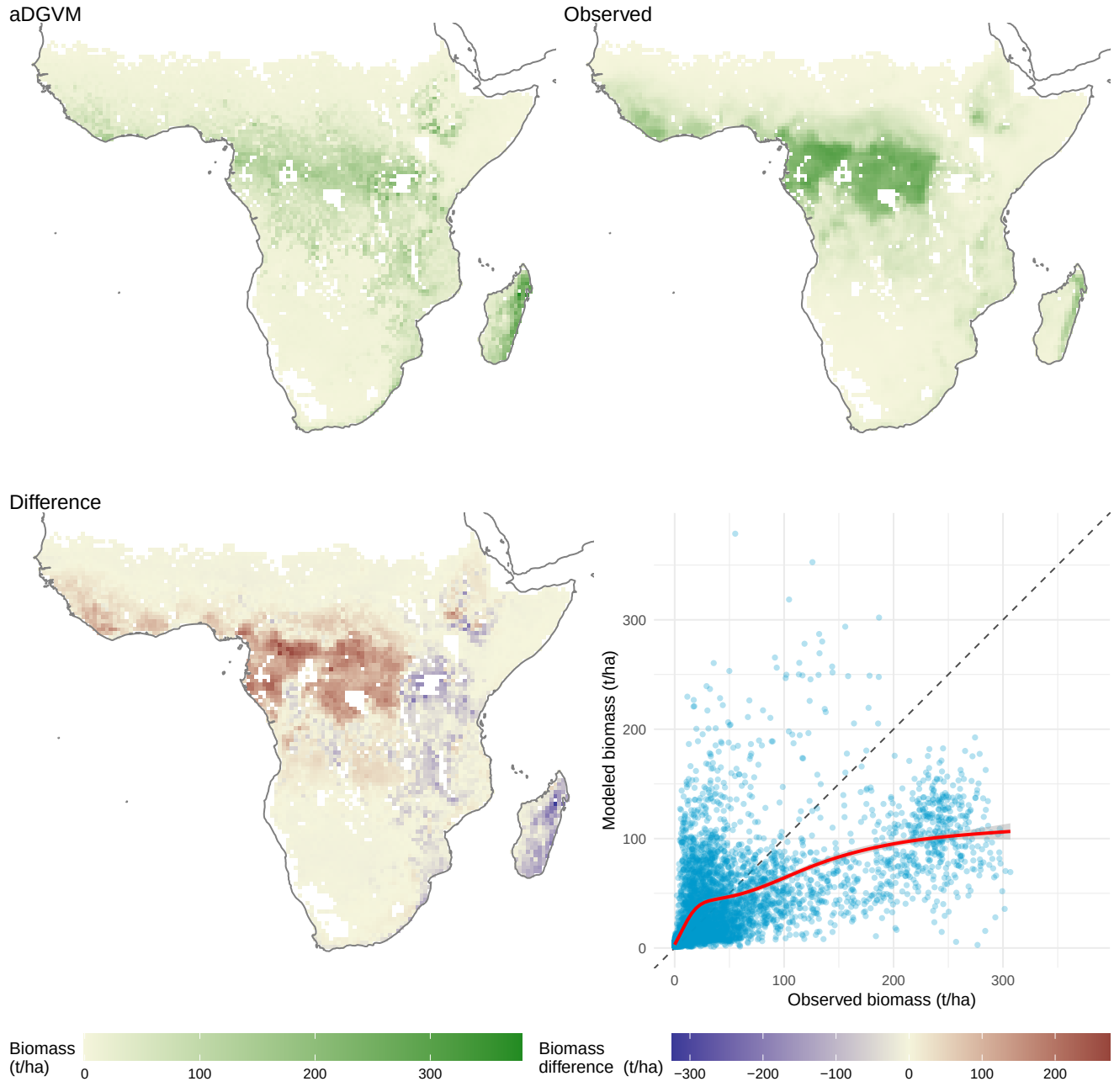

Figure 1. Modeled and observed aboveground biomass. Modeled aboveground biomass was derived from aDGVM simulations in Martens et al. (2021), observation-based biomass from Ross et al. (2021).

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

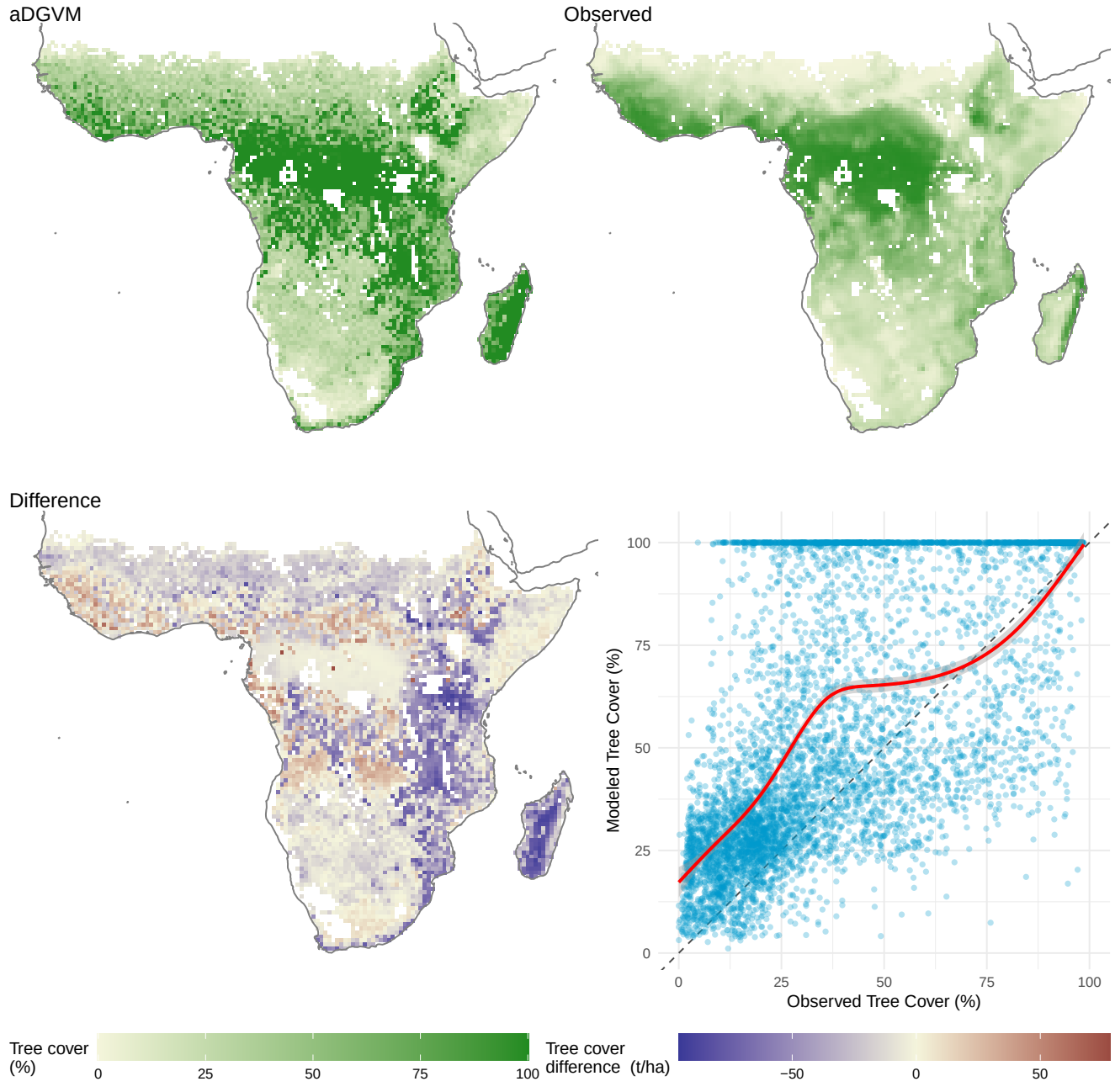

Figure 2. Modeled and observed tree cover. Modeled tree cover was derived from aDGVM simulations in Martens et al. (2021), observation-based tree cover from Ross et al. (2021).

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

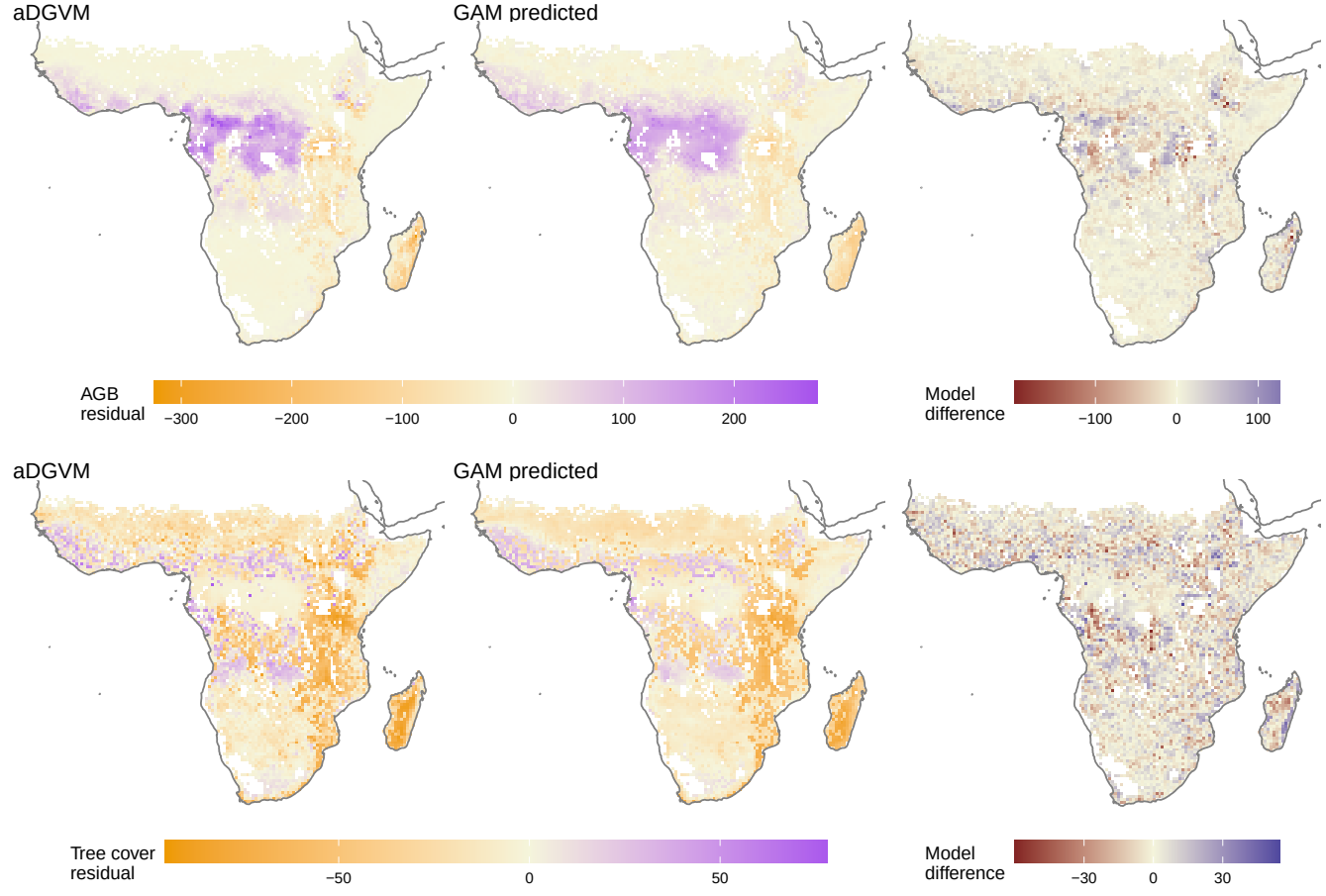

Figure 3. aDGVM and GAM residuals. Residuals of aboveground biomass and tree cover between aDGVM simulations and Ross et al. (2021) observation-based map, as well as residuals predicted by the final GAM M9 including spatial effects, biome type, bioclimatic variables, and anthropogenic factors.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

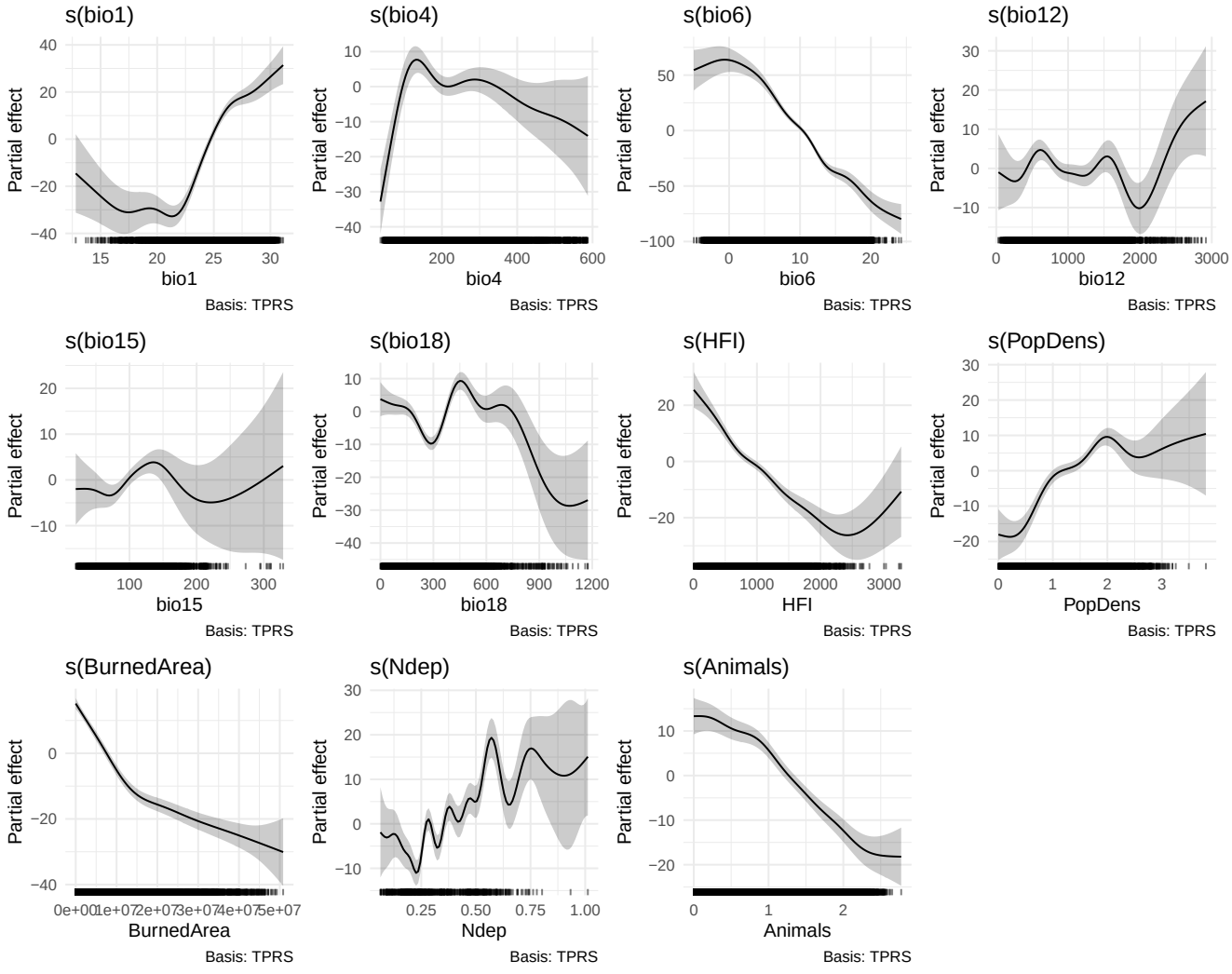

Figure 4. Partial effects in GAMs for AGB. The figure shows the partial effects of variables included in the final GAM M9 for aboveground biomass.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

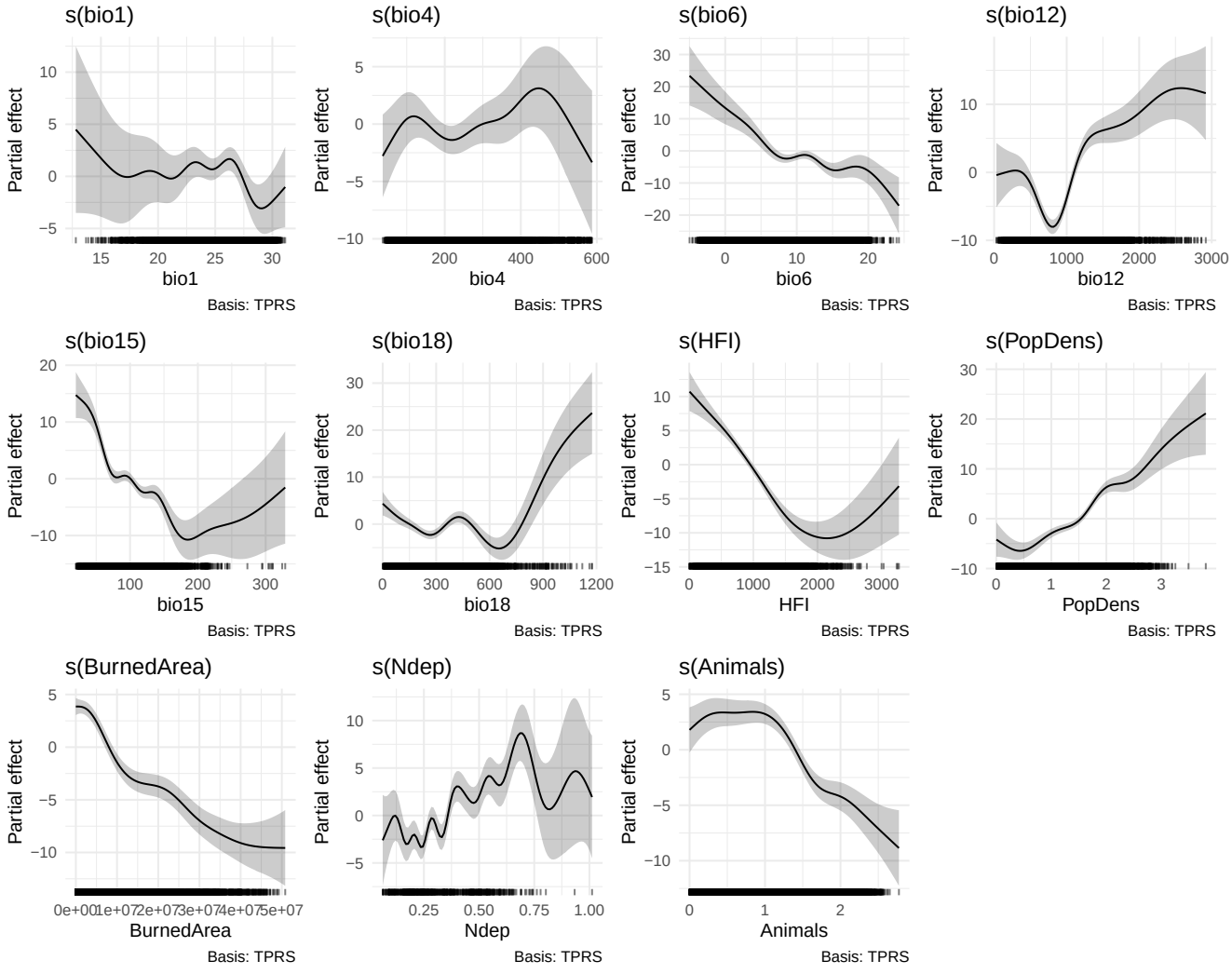

Figure 5. Partial effects in GAMs for tree cover. The figure shows the partial effects of variables included in the final GAM M9 for tree cover.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

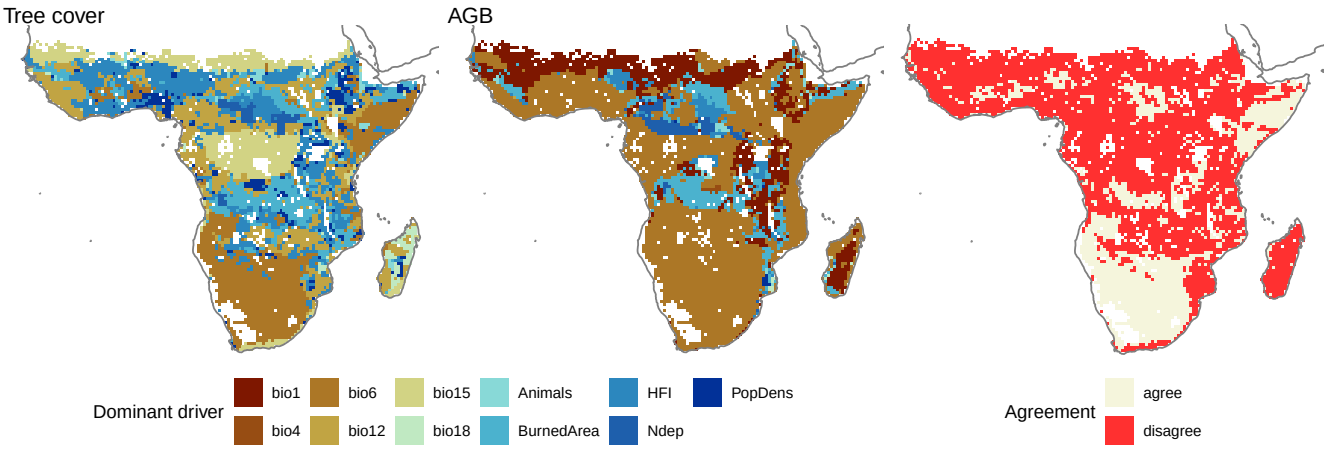

Figure 6. Dominant driver for tree cover and AGB. Drivers were obtained by identifying the predictor with the maximum absolute partial effect in the GAM predictions. Dominant drivers differ largely between tree cover and AGB.

https://doi.org/10.5194/egusphere-2026-3495
Preprint. Discussion started: 15 July 2026

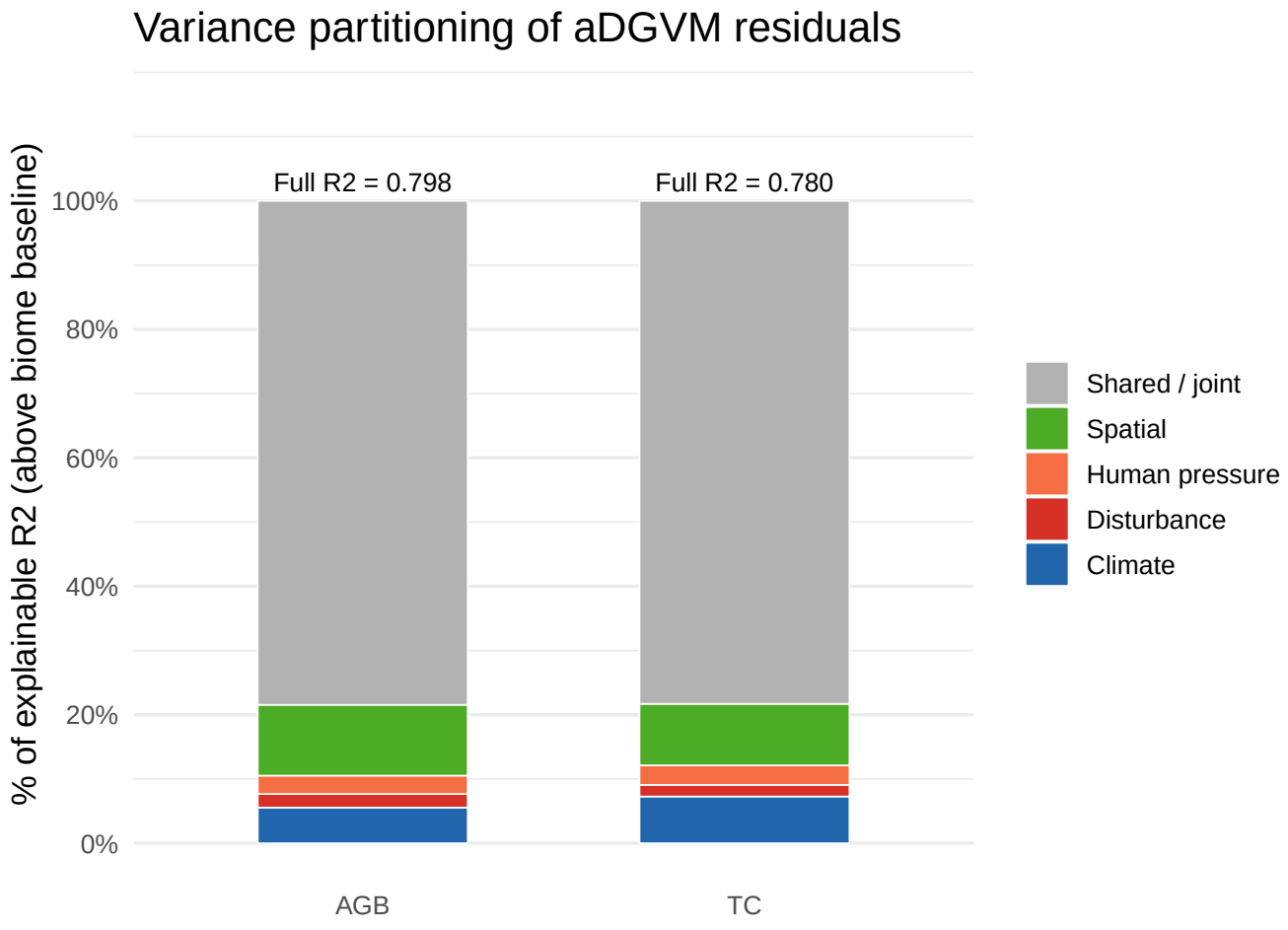

Figure 7. Variance partitioning of GAM residuals for aboveground biomass (AGB) and tree cover (TC). Bars show the percentage of total explainable R^2 (full model minus biome baseline) attributed to each component. Unique contributions are the marginal R^2 gained when each group is added last to the full model. Shared/joint variance is co-explained by two or more groups and reflects the spatial co-linearity of predictor groups along Africa's main ecological gradients. C = Climate, D = Disturbance, H = Human pressure, S = Spatial smooth.