

Attribution of aDGVM mismatches in biomass and tree cover across Africa – Supplementary materials –

Simon Scheiter^{1*}, Liam Langan¹

15th June 2026

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

*Author for correspondence:

E-mail: simon.scheiter@senckenberg.de, phone: ++49 69 7542 1864

Table S1: AIC and AIC change in forward selection of bioclimatic variables. For the analysis, each bioclimatic variables was added individually to the base model including all predictors except climate. Changes in AIC were tracked.

Variable	AGB		Tree cover	
	AIC	Δ AIC	AIC	Δ AIC
Base model	57386.73	—	48387.09	—
bio1	57279.84	−106.89	48254.29	−132.80
bio2	57176.27	−210.46	48309.24	− 77.85
bio3	57196.56	−190.17	48281.87	−105.22
bio4	57171.98	−214.75	48225.11	−161.98
bio5	57327.82	− 58.91	48171.06	−216.03
bio6	57115.05	−271.68	48156.22	−230.86
bio7	57092.05	−294.67	48241.41	−145.68
bio8	57276.74	−109.99	48265.65	−121.43
bio9	57251.74	−134.99	48260.74	−126.35
bio10	57288.89	− 97.84	48186.02	−201.07
bio11	57247.01	−139.72	48281.24	−105.85
bio12	57290.72	− 96.01	47829.52	−557.57
bio13	57315.42	− 71.31	48297.97	− 89.12
bio14	57237.51	−149.22	48300.84	− 86.25
bio15	57341.83	− 44.90	48209.37	−177.72
bio16	57331.42	− 55.31	48187.58	−199.51
bio17	57048.05	−338.68	48202.64	−184.45
bio18	57000.85	−385.88	48226.30	−160.79
bio19	57292.19	− 94.54	48242.70	−144.38

Table S2: Model formulations used for assessing predictor combinations.

Model	Formula
M0	$r_i^{var} = s(x, y, \text{bs} = "tp", k = 50) + \varepsilon_i$
M1	$r_i^{var} = \text{biome} + s(x, y, \text{bs} = "tp", k = 50) + \varepsilon_i$
M2	$r_i^{var} = s(\text{bio1}) + s(\text{bio4}) + s(\text{bio6}) + s(\text{bio12}) + s(\text{bio15}) + s(\text{bio18}) + s(x, y) + \varepsilon_i$
M3	$r_i^{var} = \text{biome} + s(\text{bio1}) + s(\text{bio4}) + s(\text{bio6}) + s(\text{bio12}) + s(\text{bio15}) + s(\text{bio18}) + s(x, y) + \varepsilon_i$
M4	$r_i^{var} = s(\text{HFI}) + s(\text{PopDens}) + s(\text{Ndep}) + s(\text{Animals}) + s(x, y) + \varepsilon_i$
M5	$r_i^{var} = \text{biome} + s(\text{HFI}) + s(\text{PopDens}) + s(\text{Ndep}) + s(\text{Animals}) + s(x, y) + \varepsilon_i$
M6	$r_i^{var} = s(\text{HFI}) + s(\text{PopDens}) + s(\text{BurnedArea}) + s(\text{Ndep}) + s(\text{Animals}) + s(x, y) + \varepsilon_i$
M7	$r_i^{var} = \text{biome} + s(\text{HFI}) + s(\text{PopDens}) + s(\text{BurnedArea}) + s(\text{Ndep}) + s(\text{Animals}) + s(x, y) + \varepsilon_i$
M8	$r_i^{var} = s(\text{bio1}) + \dots + s(\text{bio18}) + s(\text{HFI}) + s(\text{PopDens}) + s(\text{BurnedArea}) + s(\text{Ndep}) + s(\text{Animals}) + s(x, y) + \varepsilon_i$
M9	$r_i^{var} = \text{biome} + s(\text{bio1}) + \dots + s(\text{bio18}) + s(\text{HFI}) + s(\text{PopDens}) + s(\text{BurnedArea}) + s(\text{Ndep}) + s(\text{Animals}) + s(x, y) + \varepsilon_i$

Table S3: Predictor inclusion matrix for candidate models. An “X” indicates that the term is included in the model.

Model	Biome	Climate	HFI	PD	BA	Ndep	LS	Spatial
M0								X
M1	X							X
M2		X						X
M3	X	X						X
M4			X	X		X	X	X
M5	X		X	X		X	X	X
M6			X	X	X	X	X	X
M7	X		X	X	X	X	X	X
M8		X	X	X	X	X	X	X
M9	X	X	X	X	X	X	X	X

Table S4: Variance partitioning of GAM residuals. Unique contributions are the marginal R^2 gained when each predictor group is added last to the full model. Shared/joint variance is the remainder of total explainable R^2 (full model minus biome baseline) after subtracting all unique contributions. Values in parentheses give the percentage of total explainable R^2 . C = Climate, D = Disturbance, H = Human pressure, S = Spatial smooth.

Component	AGB		TC	
	R^2	%	R^2	%
Biome baseline	0.011	—	0.298	—
Full model	0.798	—	0.780	—
Total explainable	0.787	100%	0.482	100%
Unique C (Climate)	0.044	5.5%	0.035	7.3%
Unique D (Disturbance)	0.017	2.1%	0.009	1.8%
Unique H (Human)	0.023	2.9%	0.015	3.0%
Unique S (Spatial)	0.087	11.0%	0.046	9.6%
Shared / joint	0.618	78.5%	0.377	78.3%

Table S5: Smooth term statistics of the final GAM M9 for aboveground biomass (AGB) and tree cover (TC). Effective degrees of freedom (edf) reflect the complexity of each smooth; F is the test statistic for the significance of each smooth. Biome is included as a parametric factor and is not shown here. All terms were retained in the final model.

Predictor	AGB			TC		
	edf	F	p	edf	F	p
bio1 (Mean annual temperature)	6.95	31.8	< 0.0001	6.60	3.2	< 0.0001
bio4 (Temperature seasonality)	8.09	14.7	< 0.0001	5.60	2.4	< 0.001
bio6 (Min. temperature coldest month)	7.36	42.4	< 0.0001	7.71	9.5	< 0.0001
bio12 (Annual precipitation)	7.32	7.4	< 0.0001	7.37	31.8	< 0.0001
bio15 (Precipitation seasonality)	5.15	2.4	< 0.001	7.41	14.2	< 0.0001
bio18 (Precipitation warmest quarter)	7.56	20.8	< 0.0001	6.99	12.2	< 0.0001
Human Footprint Index (HFI)	5.28	16.3	< 0.0001	4.34	18.4	< 0.0001
Population density	6.57	12.1	< 0.0001	6.41	15.5	< 0.0001
Burned area	5.06	56.4	< 0.0001	5.52	23.4	< 0.0001
N deposition	15.64	7.7	< 0.0001	14.60	4.4	< 0.0001
Livestock density	4.78	14.8	< 0.0001	5.23	10.5	< 0.0001
Spatial smooth $s(x, y)$	47.95	53.9	< 0.0001	46.12	25.7	< 0.0001

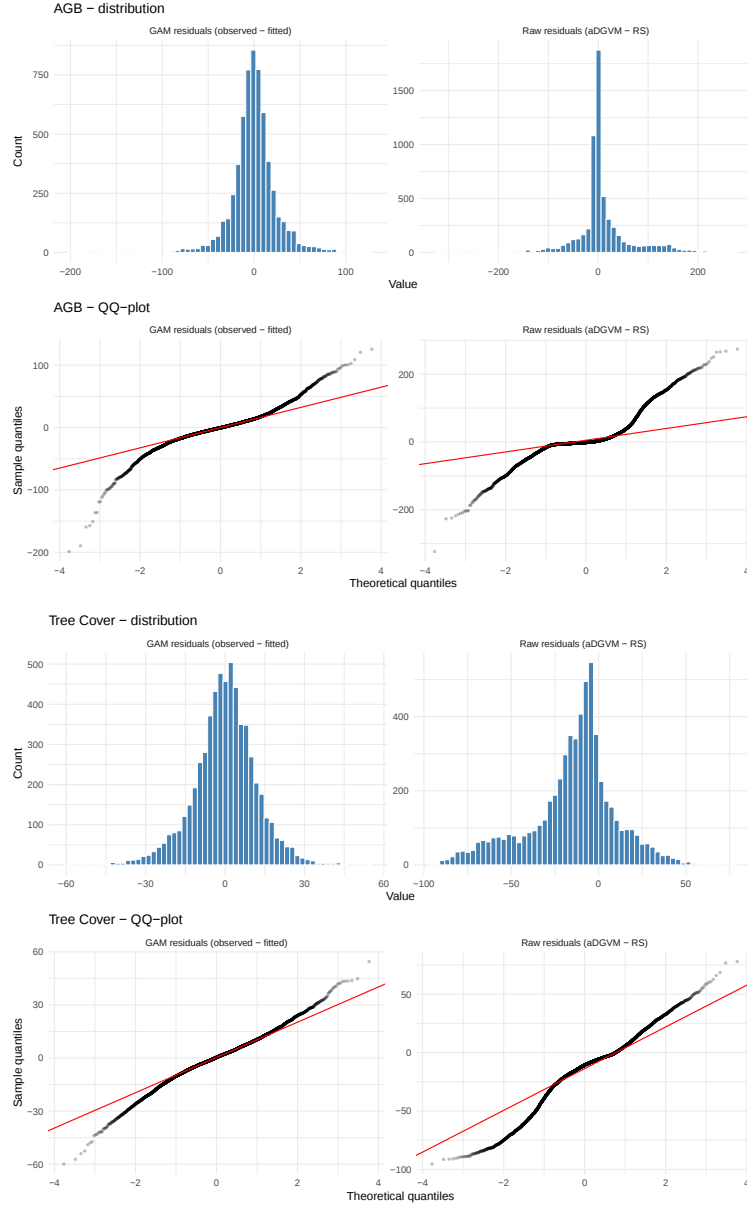


Figure S1: Distribution of raw residuals (aDGVM minus remote sensing) and GAM residuals (observed minus fitted) for aboveground biomass (top) and tree cover (bottom). Left panels show histograms, right panels show quantile-quantile plots against a theoretical Gaussian distribution (red line). Raw AGB residuals are moderately right-skewed (skewness = 0.89); raw TC residuals are more symmetric (skewness = -0.51). After fitting GAM M9, both distributions are near-symmetric, supporting the use of a Gaussian error family.

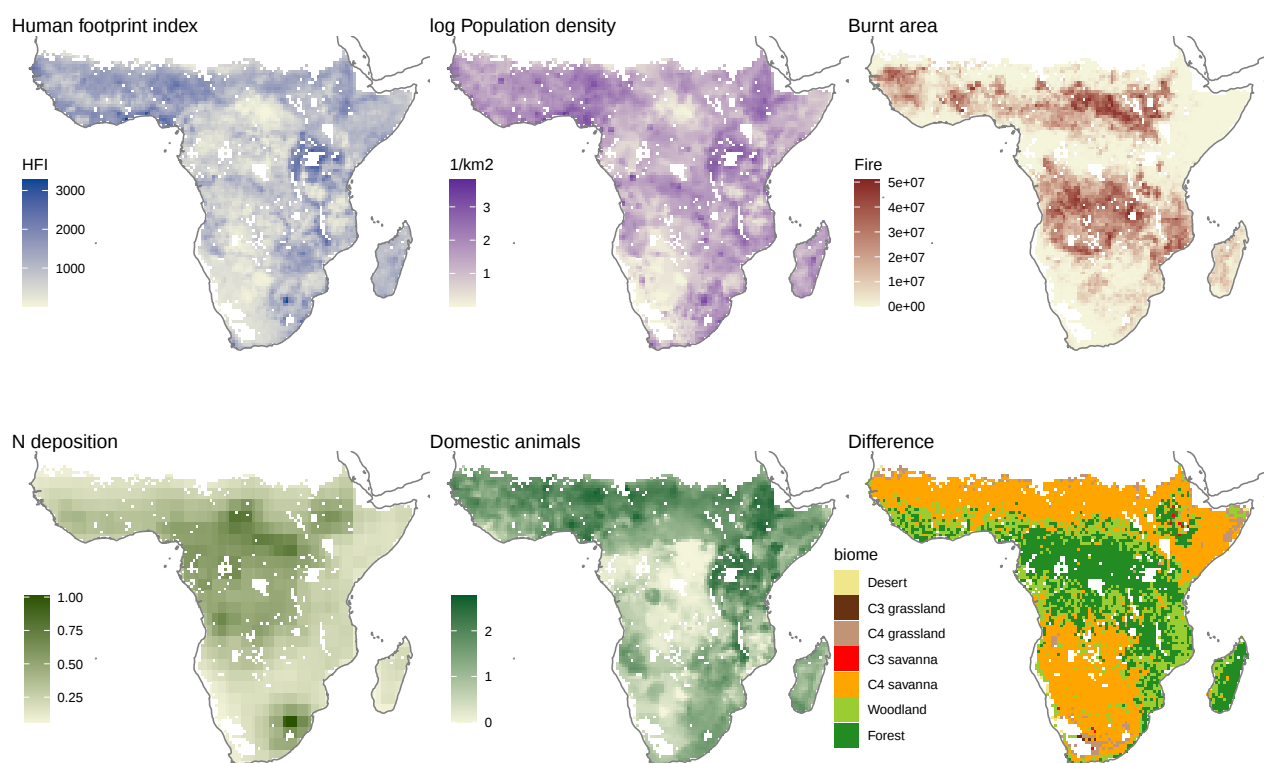


Figure S2: Predictors used in the GAMs. Observation-based variables used in the GAMs. References for variables are provided in section Materials and Methods.