Article

Why models underestimate West African tropical forest primary productivity

Huanyuan Zhang-Zheng [1,2] ✉, Xiongjie Deng [1], Jesús Aguirre-Gutiérrez [1,2], Benjamin D. Stocker [3,4], Eleanor Thomson[1], Ruijie Ding[5], Stephen Adu-Bredu[6,7], Akwasi Duah-Gyamfi [6], Agne Gvozdevaite[1], Sam Moore[1], Imma Oliveras Menor [1,8], I. Colin Prentice [5] ✉ & Yadvinder Malhi [1,2] ✉

Tropical forests dominate terrestrial photosynthesis, yet there are major contradictions in our understanding due to a lack of field studies, especially outside the tropical Americas. A recent field study indicated that West African forests have among the highest forests gross primary productivity (GPP) yet observed, contradicting models that rank them lower than Amazonian forests. Here, we show possible reasons for this data-model mismatch. We found that biometric GPP measurements are on average 56.3% higher than multiple global GPP products at the study sites. The underestimation of GPP largely disappears when a standard photosynthesis model is informed by local field-measured values of (a) fractional absorbed photosynthetic radiation (fAPAR), and (b) photosynthetic traits. Remote sensing products systematically underestimate fAPAR (33.9% on average at study sites) due to cloud contamination issues. The study highlights the potential widespread underestimation of tropical forests GPP and carbon cycling and hints at the ways forward for model and input data improvement.

Carbon exchanges between terrestrial ecosystems, especially tropical forests, and the atmosphere are a major element of the global carbon cycle. As the world's most productive terrestrial ecosystems[1], tropical forests have been estimated to account for around 44% of global forest biomass and 43% of global gross primary production (GPP)[2,3]. Nonetheless, confidence in estimates of tropical forest productivity remains low[4,5]. There is still a large variation among models regarding the magnitude and spatial pattern of tropical GPP[6–8]. Multiple global-scale forest GPP studies have indicated that tropical forests have larger uncertainty compared to other biomes[9–12]. This is likely an inevitable consequence of the paucity of carbon cycling data in the tropics relative to temperate regions[13–16]. For instance, many previous studies

found particularly large GPP data-model discrepancies for tropical forests between flux towers and models (including satellite GPP products)[16–19], but investigation into the causes of discrepancy was not attempted due to the lack of in situ auxiliary measurements of plant traits and forest characteristics[20]. The large unresolved data-model discrepancy suggests fundamental challenges in our understanding of tropical forest productivity and its geography.

The uncertainty in estimates of tropical forest productivity is particularly large in West Africa[21]. The first field quantification of GPP in African forests (termed 'biometric GPP') reported possibly the highest GPP value recorded in intact forests, which however, are underestimated by about 60% in both the MODIS and FLUXCOM GPP

[1]Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, United Kingdom. [2]Leverhulme Centre for Nature Recovery, University of Oxford, Oxford, UK. [3]Institute of Geography, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland. [4]Oeschger Centre for Climate Change Research, University of Bern, Falkenplatz 16, 3012 Bern, Switzerland. [5]Georgina Mace Centre for the Living Planet, Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK. [6]Forestry Research Institute of Ghana, Council for Scientific and Industrial Research, Kumasi, Ghana. [7]Department of Natural Resources Management, CSIR College of Science and Technology, Kumasi, Ghana. [8]AMAP (Botanique et Modelisation de l'Architecture des Plantes et des Végétations), CIRAD, CNRS, INRA, IRD, Université de Montpellier, Montpellier, France. ✉e-mail: huanyuan.zhang@ouce.ox.ac.uk; c.prentice@imperial.ac.uk; yadvinder.malhi@ouce.ox.ac.uk

products, two widely used global maps of tropical forest productivity[22]. Since GPP is the source of carbon and energy for terrestrial ecosystems in Dynamic global vegetation models (DGVMs)[23], the data-model discrepancy in GPP propagates into NPP (Fig. S5) and possibly other downstream variables. It has not yet been explained, from modellers' perspective, why this area has a GPP higher than simulations. As models are designed based on current ecological theory[24], such a large discrepancy of multiple sites signals a lack of physiological understanding or poor parameterisation of forest physiology.

In this work, we set out to explain the high productivity of West African forests using photosynthesis models and investigate the reasons behind the data-model discrepancy, with a view to provide information to modellers working on assessing and mapping the productivity of forests globally. The investigation follows three objectives (Table 1). Objective (1): We first compare the biometric GPP to multiple models and satellite-based products, to quantify the data-model discrepancy. Objective (2): We investigate whether the field-observed high productivity is consistent with the leaf photosynthetic traits and other field measurements that were commonly simulated by models. This investigation is made possible by extensive field measurements of environmental variables, plant traits, and carbon fluxes at the study sites[25–28]. Objective (3): We attempt to account for the data-model discrepancy of MODIS and Pmodel (Table 2) and explore key responsible parameters in DGVMs. This is done by substituting model parameters with field measurements, according to several hypotheses listed below as to the cause of this discrepancy.

Most vegetation models and global GPP products calculate GPP using key inputs including (1) light use efficiency (LUE), a key parameter of MODIS-GPP and Pmodel. LUE could not be directly measured but could be derived from field-measured photosynthetic capacity[29] (Supplementary method). Pmodel[29] and most DGVMs use photosynthetic capacity to calculate GPP[30]; (2) fraction of absorbed photosynthetically active radiation (fAPAR), a key parameter of MODIS-GPP and Pmodel. Models typically simulate LAI and then express fAPAR as a function of LAI and canopy extinction coefficient; (3) plant functional types (PFTs) classification; (4) climate variables, including temperature, relative humidity and photosynthetic photon flux density (PPFD), often calculated from incoming shortwave radiation. For models these are retrieved from global climate data products.

Therefore, for objective (3), we hypothesise that the large data-model discrepancy for GPP could stem from inaccuracies in one or more of these key input variables: (Hypothesis 1) incorrect LUE, or photosynthetic capacity; (Hypothesis 2) incorrect fAPAR; or (Hypothesis 3) inappropriate assignment of plants functional types; or (Hypothesis 4) climate variables. Multiple experiments are designed to test each hypothesis (Table 1, see Methods for detailed experiment settings).

Overall, we identify key areas for model improvement and also provide a trait-based explanation for the high GPP observed in West Africa.

Results

Intercomparison of GPP estimates
Each study site represents a major forest type of West African forests (Fig. 1). For all study sites, considerable discrepancies in GPP are found among DGVMs, FLUXCOM, MODIS, Pmodel, and biometric measurements, with a pattern that is consistent from site to site (Fig. 2). At each site, biometric measured GPP exceeds methods GPP and global GPP products. Overall, the DGVMs' average GPP is higher than FLUXCOM, which in turn is higher than MODIS GPP. The discrepancy between biometric and MODIS GPP is about 20 MgC/ha/year regardless of sites. The largest data-model discrepancy between DGVMs and biometric GPP is seen at BOB (22.8 versus 43.3 MgC/ha/year). At BOB, the DGVM average, FLUXCOM, and MODIS report almost equal GPP – but all less than the measured GPP. There are substantial disagreements between different DGVMs; for example, at ANK, the DGVMs simulated GPP ranged from 14.3 to 42.6 (biometric GPP at 40.1) MgC/ha/year.

The match between trait-based GPP and biometric GPP
Using field observed PPFD, fAPAR, and photosynthetic traits as input, Pmodel could estimate a GPP (i.e., experiment Pmodel_PfL) slightly lower but still within the uncertainty range of field biometric GPP at BOB and KOG sites. At the ANK site, Pmodel_PfL greatly reduces the data-model discrepancy but remains lower than biometric GPP (Fig. 3). Nonetheless, the slight mismatch at ANK does not undermine further investigation at ANK because there are more dominant factors contributing to the data-model discrepancy, explained below.

Photosynthetic trait and fAPAR
It was hypothesised that GPP data-model discrepancy could stem from inaccuracies in (Hypothesis 1) LUE, or photosynthetic capacity; (Hypothesis 2) fAPAR. Photosynthetic capacity estimated by the Pmodel is lower, although close to field measured photosynthetic capacity (Fig. S2) implying underestimated LUE. As a consequence, Pmodel_Pf is lower than Pmodel_PfL consistently at all sites (Fig. 3).

GPP is substantially underestimated when in situ fAPAR is replaced with MODIS fAPAR, resulting in a considerable difference between Pmodel_Pf GPP and Pmodel_P at any site. Especially at ANK, fAPAR is found to be the largest contributor to GPP data-model discrepancy. A further investigation into MODIS fAPAR (Fig. 4) shows that during the rainy season, MODIS fAPAR decreases along with increasing cloud cover (up to 90% of pixels), whereas field fAPAR remains fairly constant during the year. Filtering fAPAR values according to cloudiness or gap-filling of bad data points would remove most of the values from the rainy season, leading to an underestimated mean annual fAPAR. The above issue affects a major proportion of the African tropical forest, and likely many areas of the tropics (Fig. 1).

By comparing to Pmodel_null, we showed that GMAO PPFD lead to overestimation of GPP at site ANK and BOB, but leads to slight underestimation of GPP at site KOG (Fig. S6). This is associated with the low resolution of GMAO PPFD. Using a higher resolution PPFD product (e.g. ERA5, Fig. S3) could alleviate this issue. Thus, we

Table 1 | Study objectives, hypotheses and testing datasets

Objectives	Data	Figure
1, quantify the data-model discrepancy	Pmodel_null, MODIS, Biometric, FLUXCOM, TRENDY	Fig. 2
2, consistency between leaf photosynthetic traits and biometric GPP	Compare Pmodel_PfL to Biometric	Fig. 3
3, Possible sources of data-model discrepancy (including 4 hypotheses)		
Hypothesis 1 LUE and photosynthetic capacity	Compare Pmodel_PfL to Pmodel_Pf; Compare Pmodel_null to MODIS	Fig. 3; Figure S6
Hypothesis 2 fAPAR	Compare Pmodel_Pf to Pmodel_P	Fig. 3
Hypothesis 3 Plant functional types	Compare TRENDY Forest-only GPP to whole-pixel GPP	Fig. S4
Hypothesis 4 Climate variables	Compare ERA5-Land, MERRA-2 (by GMAO), CRU-JRA to field measurements	Fig. S3

Table 2 | Sources of gross primary production (GPP)

GPP	Main features	Group
Biometric	Measuring Each GPP component (e.g. leaf NPP, stem respiration) in the field.	Field works
TRENDY	Global dynamic vegetation models that used environmental variables (but not fAPAR) as inputs.	Model
Pmodel	a light use efficiency model, which was used to design multiple experiments to test hypotheses for Objective 3. The experiments include Pmodel_null, Pmodel_PfL, Pmodel_Pf, and Pmodel_P (see Fig. 3).	Model
MODIS	MODIS GPP is a light use efficiency model that calculates GPP using satellite observed fAPAR.	Observation based products
FLUXCOM	By upscaling carbon fluxes observed by eddy covariance towers.	Observation based products

conclude that PPFD could not explain the underestimation of GPP at the study sites. At the ANK site, MODIS and Pmodel_P GPP are almost identical (Fig. 3a), but at BOB and KOG, MODIS is lower than Pmodel_P GPP. This portion of data-model discrepancy could be resolved by using the optimality-based LUE other than MODIS LUE (from the look-up table) (Fig. S6). This portion of data-model discrepancy could not be resolved by PPFD as explained above. Although MODIS LUE is not publicly available, the comparison between Pmodel_null and MODIS suggests that optimality theory tends to estimate an LUE higher than MODIS LUE. In short, Hypotheses 1 and 2 are accepted.

Scaling and plant functional type

Hypothesis 3 suggests that the GPP data-model discrepancy originated from inappropriate assignments of plant functional types. At BOB the entire forest fragment is smaller than a single grid cell in a DGVM but is hundreds of times larger than a MODIS grid cell (scales illustrated in Fig. S1). MODIS GPP estimates that the surrounding area (a composite of savanna and cocoa farms) has only half the productivity of the forests. Therefore, the mean of the whole map (19.6 MgC/ha/year) appears to be smaller than that of the forests (23 MgC/ha/year). This is not the case for TRENDY, which simulated a comparable value between forests-only and whole grid cell GPP (Fig. S4) (i.e. similar simulated GPP between forests and C4 grass). There are no field studies to our knowledge comparing cocoa farm GPP to adjacent forests GPP.

For all the study sites, most TRENDY models correctly simulate the grid cell as a composite of 'forest' and 'C4 grass', although the fraction of forests varies considerably among models (Table S1). Given that the 'forest-only' GPP of TRENDY and MODIS are much lower than biometric GPP (Figs. 2; S4; S1) we conclude that the GPP data-model discrepancy between TRENDY and biometric GPP cannot be explained by scaling or PFT issues (rejecting hypothesis 3). Besides, please note that TRENDY S2 keeps the land cover state fixed to its pre-industrial state of 1700 and thus the proportion of C4 grass (Table S1) is temporarily invariant and could be smaller than a land cover dynamic simulation (e.g. MODIS replies on satellite imageries that incorporates real-time land cover change).

Data-model consistency in climate variables

Hypothesis 4 suggests that the GPP data-model discrepancy originated from climate variables. The data-model differences we found for temperature, relative humidity, PPFD and c_i/c_a were too small to explain the GPP data-model discrepancy (Fig. S3), thus rejecting hypothesis 4.

To sum up, the data-model discrepancy for GPP at the study sites mainly stems from: incorrect LUE, or photosynthetic capacity (Hypothesis 1 accepted); and incorrect fAPAR (Hypothesis 2 accepted); not inappropriate assignment of plants functional types (Hypothesis 3 rejected); nor climate variables (Hypothesis 4 rejected).

Discussion

Main sources of GPP data-model discrepancy

We not only demonstrate a significant data-model discrepancy in West African forest GPP but also reveal the main sources of this discrepancy.

At the Ankasa wet rainforest site, we found that underestimation of fAPAR, likely because of cloud contamination, is the major contributor to GPP data-model discrepancy. At the Bobiri (semideciduous) and Kogyae (dry forest) sites, too low values of photosynthetic traits (leading to a bias in light use efficiency) are the primary source of GPP data-model discrepancy, which can be partly explained using optimality theory predictions. Additionally, fAPAR accounts for a large proportion of GPP data-model discrepancy at KOG but not at BOB. Through this analysis we are able to fully account for the data-model discrepancy at KOG and BOB, but leave part of the data-model discrepancy unresolved (20.5% of biometric GPP) at ANK, which may be caused by inaccurate traits measurements, photosynthesis model assumption (e.g. the big leaf assumption) or possible bias in field measurements of the biometric GPP (e.g. the upscaling of leaf and stem respiration; the estimation of roots exudation)[31].

Why do models underestimate GPP in West African tropical forests?

Focusing on three sites (comprising 14 one-hectare plots) in Ghana, we have utilised a recent in situ quantification of African forest GPP and conducted a systematic data-model comparison. Despite the uncertainty associated with biometric GPP measurements, it is more likely that FLUXCOM, MODIS, TRENDY models, and Pmodel_null underestimate West African forest productivity than that the biometric GPP measurements overestimate it. This is because (1) the biometric GPP is an average across multiple plots spanning several years. To our knowledge, there are no other biometric GPP measurements in the study region, but the high GPP is supported by another field study reporting very high forests biometric NPP[32]; (2) from a photosynthesis traits perspective, West African seasonally dry forests are characterised by a high CO_2 assimilation rate or high photosynthetic capacity, in comparison to wet-evergreen tropical forests studied in other continents, predominately in Amazonia[27,33,34]; Pmodel could simulate a GPP close to biometric GPP if informed by field-measured traits (Fig. 3), which signals a broad consistency between the high GPP and observed photosynthetic traits. (3) the cause of the models' underestimation can be well explained by errors in modelled LUE and satellite-derived fAPAR (Fig. 3). (4) There is only one flux tower for West African forests (site GH-ank at the Ankasa site), which reports three years of GPP varying substantially[19]. The tower reported 2011 annual GPP at 36.06 MgC/ha/year, but the tower only operated for half-year in 2012 and 2014 with estimated annual GPP at 22.02 and 26.1 respectively (FLUXNET2015[35], variable GPP_NT_CUT_REF) (Fig. S8). In comparison, the biometric GPP is 40.11 and Pmodel_PfL is 31.85. Therefore, the 'true' GPP of site ANK appears to be more likely to situate around 36 MgC/ha/year with biometric GPP overestimating, and the 'unresolved data-model discrepancy' at site ANK (Fig. 3) is probably due to bias in the biometric GPP. Many tropical flux towers face logical or technical issues[36,37], and challenging subcanopy CO_2 storage estimation[38,39]. For example, the road leading to site ANK is seasonally inundated (Photos in Supplementary information) when the study site can only be accessed on foot. During FLUXCOM extrapolation, the machine learning algorithm would not receive information about BOB's higher photosynthetic capacity and would simply predict
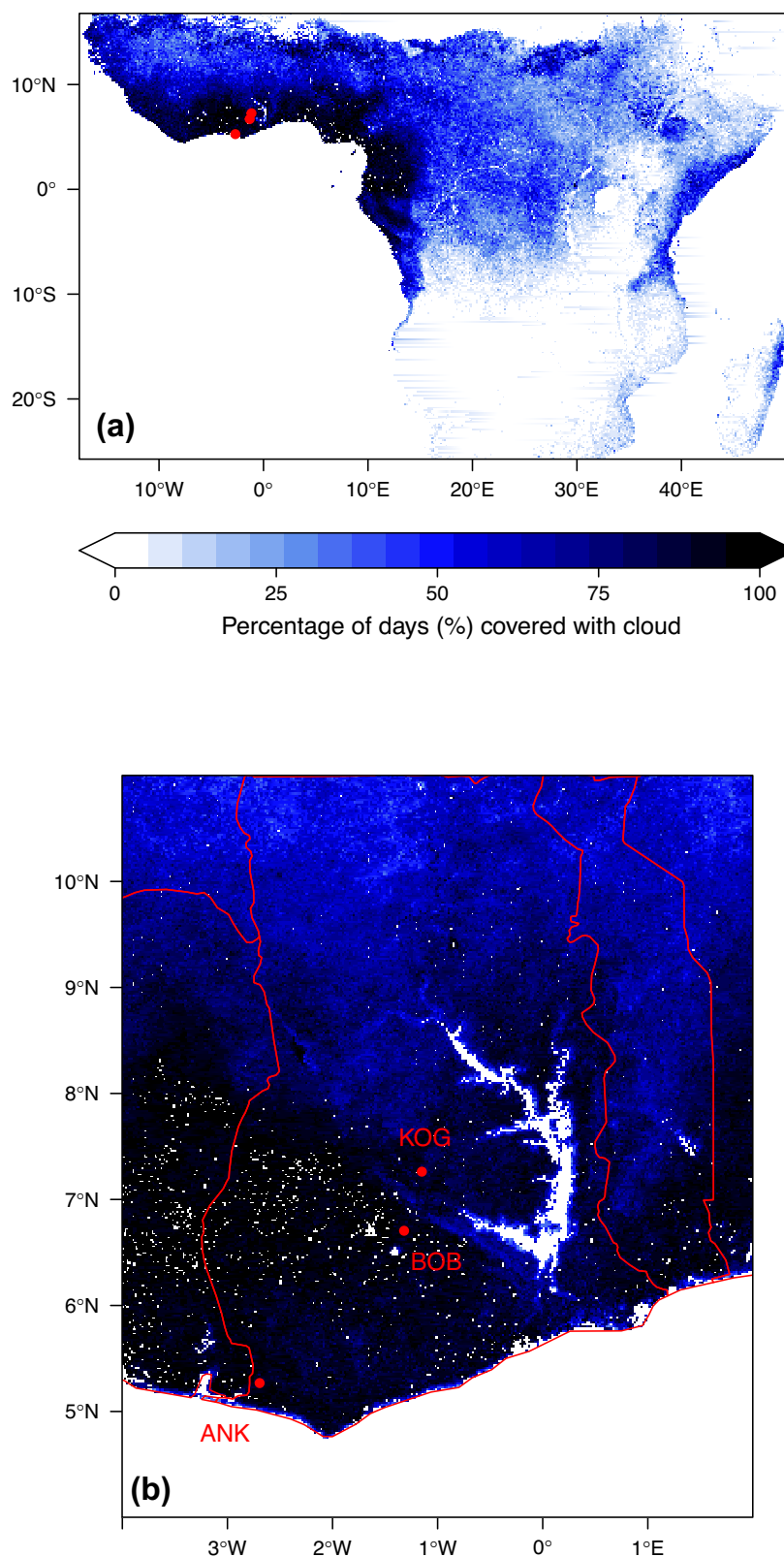
Fig. 1 | Map of the three study sites in West Africa. Panel (**a**) displays a regional view of Africa, and panel (**b**) zooms in on Ghana. Blue scales show the percentage of days contaminated by clouds during the cloudy season (August and September), which is the percentage of MODIS data marked as '01 significant clouds present' and '10 mixed clouds present'). Each red dot denotes a site. Each site contains multiple one-hectare plots.
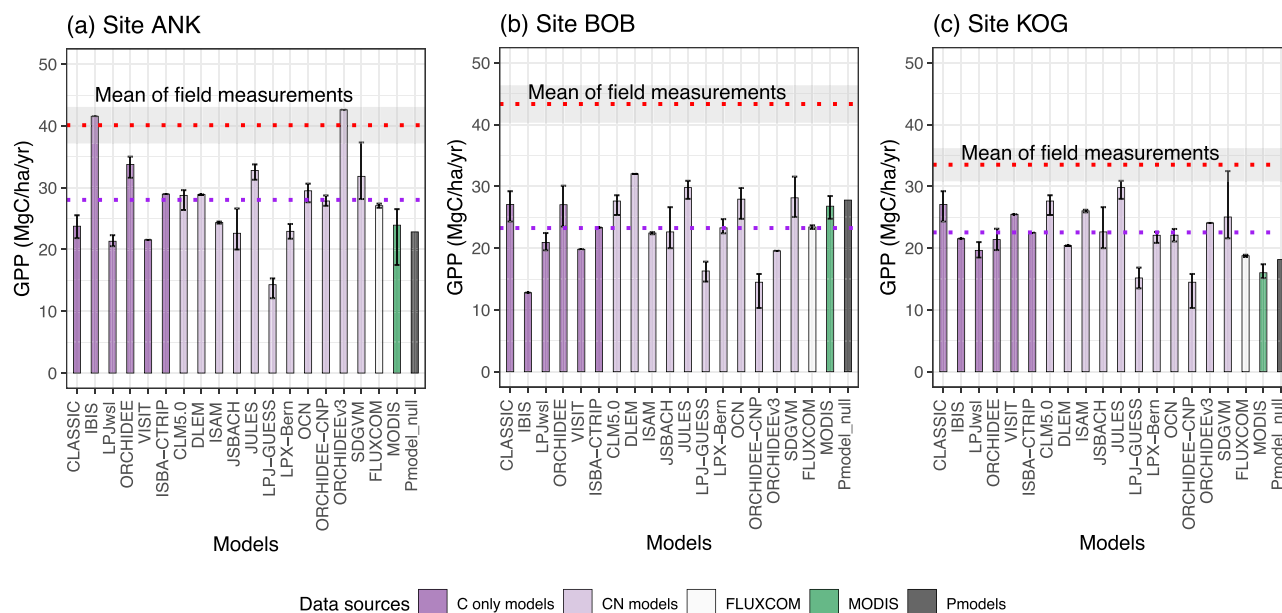
Fig. 2 | Intercomparison of gross primary productivity (GPP, MgC/ha/year) from various independent sources. The figure contains study sites (**a**) Ankasa, (**b**) Bobiri and (**c**) Kogaye. The red dotted lines denote in situ biometric GPP, as a mean of multiple one-hectare plots (Table S3); measurements are taken spanning 2011 to 2016[22,25]. The grey areas denote measurement uncertainty, not standard error. The uncertainty is calculated through error propagation. For GPP models and products, we calculate mean annual GPP from 2011 to 2016. Bars denote Carbon-only models (dark purple), Carbon-Nitrogen coupled models (light purple), FLUXCOM (white) and MODIS (green). The error bars are maximum/minimum mean annual GPP during the study period. The purple dotted line denotes all TRENDY models (purple bar) average. For TRENDY models that reported GPP per plant functional types, we display forests-only GPP (the potential GPP if the grid cell is full of forests). Otherwise, we display whole gridcell GPP. See Fig. S4 for a comparison between forest-only GPP and whole gridcell GPP. Source data are provided as a Source Data file.

BOB is less productive than ANK (opposite the observed GPP pattern) due to the lower precipitation received at BOB (Table S3).

Some degree of model underestimation of GPP may even be a pan-tropical feature. Intercomparison of GPP and NPP in previous studies[40,41] has also revealed data-model discrepancies for Amazonian lowland forests. In this study, we found that the data-model discrepancy was due to fAPAR and photosynthetic capacity. The cloudiness issue, leading to low satellite-based fAPAR, has been described as a pan-tropical phenomenon for many remotely sensed products[42–45]. Moreover, cloud-free fAPAR (Fig. 4) also underestimate fAPAR thus there are issues beyond cloud contamination. Bias in MODIS fAPAR (or LAI) will inevitably cascade into models or analyses that use MODIS fAPAR as a predictor, including FLUXCOM. Large GPP discrepancy between global flux towers and MODIS are reported, which was found associated with fAPAR and LUE not cliamte variables[19]. Many tropical rainforests have high persistent cloud cover compared to other forests and therefore may be disproportionally affected by cloud contamination[43,45].

The LUE (or photosynthetic capacity) discrepancy is strong and consistent with previous literature, since a previous study found higher A_{sat} and A_{max} values in West African species than in Amazonian ones[34]. Higher A_{sat} and A_{max} are found in BOB and KOG (drier sites), but not ANK (the wet evergreen forest on poor soils more similar to much of lowland Amazonia)[22,27], making LUE the dominating source of GPP bias in BOB and KOG but less dominant in ANK. The high photosynthetic capacity in semi-arid forest or savanna is consistent with other field studies in West Africa[33]. V_{cmax} inferred from remotely sensed leaf chlorophyll data and V_{cmax} predicted by the P model both show exceptionally high values in West Africa and parts of India – substantially higher than in Amazonia or SE Asia[46,47]. These seasonal forests on more fertile soils may have photosynthesis rates optimised to high light, temperature and VPD (also see field study[48]). This substantial spatial variability of V_{cmax} has not been incorporated in most

TRENDY models[30,49] which could lead to underestimation of GPP at such regions. Traditionally, V_{cmax} is a PFT-specific constant. If the true V_{cmax} differs among sites due to nutrient availability, then models can only embody the dynamics with varying PFT fractions. Some models do simulate V_{cmax} as a function of leaf nitrogen, in which case they could, at least in theory, reproduce those patterns. Nonetheless, leaf nitrogen may be simulated also with PFT-specific parameters and the N cycle has its own set of challenges[30]. Thus, further modelling study is needed to carefully consider the high V_{cmax} at this region.

We note that V_{cmax} behind TRENDY simulation is not directly retrievable and thus we could not conduct direct data-model comparison for V_{cmax} but discuss and compare to the PFT-specific values recalculated following models' equations[50] (Table S2). As shown in Table S2, two TRENDY models (OCN and ORCHIDEE) indeed underestimate V_{cmax}. IBIS substantially overestimates V_{cmax} which results in its high GPP at site ANK. However, JULES, CLASSIC, CLM5.0 and JSBACH have similar or higher V_{cmax} but still underestimate GPP, which implies that there are other factors contributing to the GPP underestimation. Moreover, the substantial inter-model disagreement shown in Table S2 is alarming, with V_{cmax} for evergreen forest vary from 18 to 163 µmol m^{-2} s^{-1} while field measurements at our study sites are around 30 µmol m^{-2} s^{-1}[51]. Model parameterisations of V_{cmax} for C4 grass also vary substantially and appear to be lower or higher than forests depending on models, which may affect some models that have C4 grass as main plant functional types at the study sites, for example, CLM5.0. Further, although V_{cmax} model parameters may differ from measurements at the study sites, the values may still be appropriate at a larger scale, as in-situ V_{cmax} measurement values differ considerably among sites. The wide range of V_{cmax} actually reflects substantial variation in global traits databases[46], where further data-model benchmarking study is suggested. Most TRENDY models embedded an exponential equation linking fAPAR with LAI. In other words, GPP is positively influenced by V_{cmax} and LAI. At study site KOG, most models
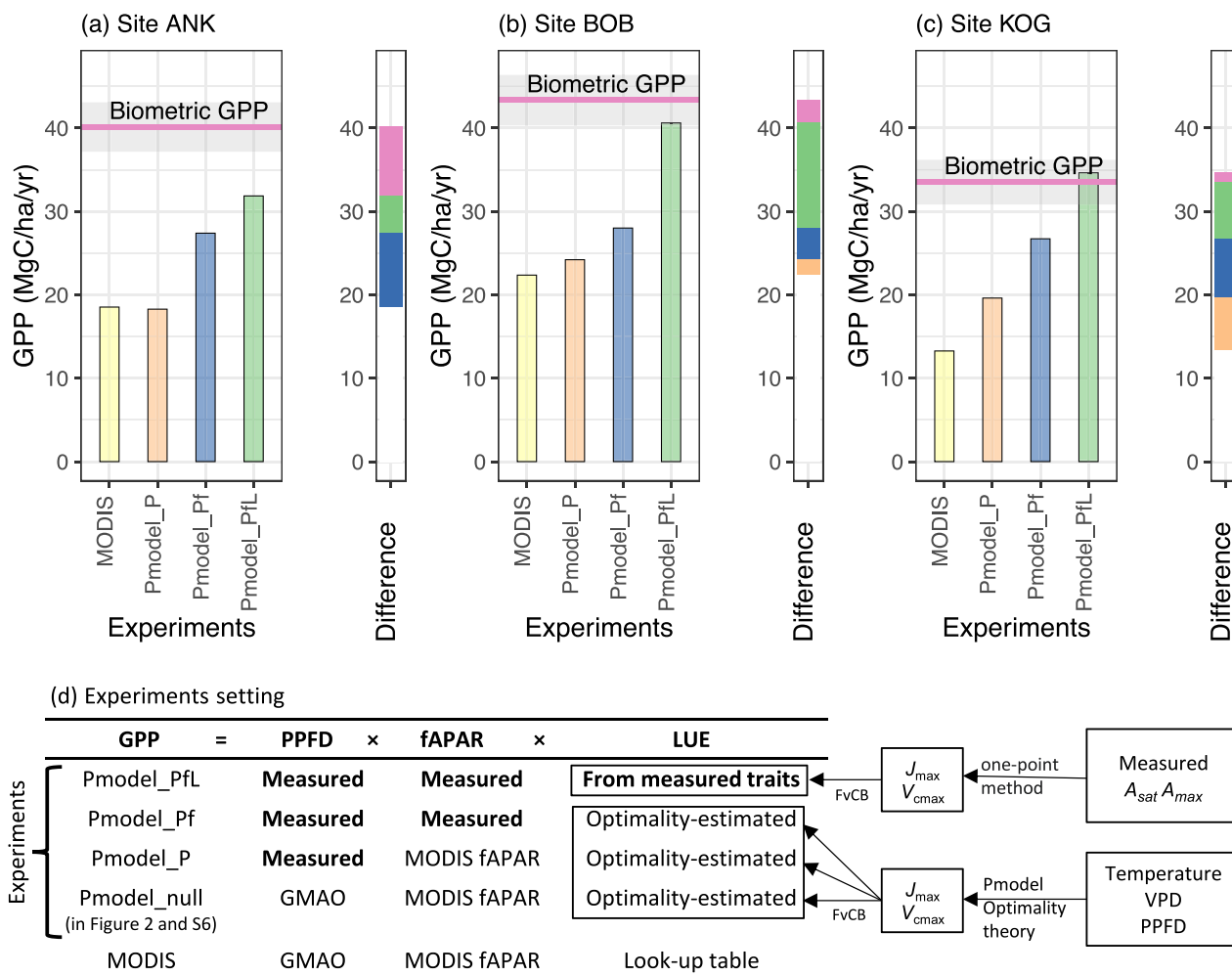
Fig. 3 | Partitioning of GPP data-model discrepancy, by comparing different experiments. Results are shown for study sites (**a**) Ankasa, (**b**) Bobiri and (**c**) Kogaye. Experiments design is shown in (**d**). The left panel displays field-based biometric GPP (pink line with grey zone showing uncertainty) and multiple GPP experiments (bars). Pmodel_PfL, Pmodel_Pf and Pmodel_P are GPP experiments which are all simulated using Pmodel but with different inputs. Both Pmodel and MODIS GPP were calculated from Eq. (1), enabling direct comparison. The right panel, a 'Diff' bar, illustrates the difference in GPP between experiments, which represents the sources of GPP data-model discrepancy. For example, the GPP difference between Pmodel_Pf and Pmodel_P experiments is caused solely by the difference in input variables - fAPAR. Therefore, the difference between them (blue) is GPP data-model discrepancy resolved by fAPAR. The difference (orange) between Pmodel_P and MODIS is GPP data-model discrepancy resolved by optimality-estimated LUE and PPFD (but mostly LUE, Fig. S6). The difference (green) between Pmodel_PfL and Pmodel_Pf is GPP data-model discrepancy caused by the difference between measured LUE and optimality-based LUE. MODIS fAPAR and GPP are from MOD15A2H and MOD17A2H excluding data marked as '01-Significant clouds were present'. Source data are provided as a Source Data file.

overestimate LAI (Figure S7). Thus, LAI would not explain the under-estimation of GPP at this site. At study sites BOB and ANK however, models averaged LAI is very close to field measurements although with large inter-model variation, suggesting that the key parameters accounting for GPP underestimation here should vary among models. To sum up, the investigation into the cause of DGVMs underestimating GPP is not straightforward. The parameterisation of V_{cmax} and LAI is one of the reasons for some models but there are likely other contributing factors that require future studies.

The importance of field evidence in productivity estimation
The availability of comprehensive field measurements allows us to trace and quantify the sources of data-model discrepancy in the GPP of

West African forests. We find that fAPAR and LUE (or photosynthetic capacities) are the dominant sources of data-model discrepancy, rather than model structure or climate variables. These findings reflect the lack of field measurements of West African forest photosynthetic traits and leaf area index (or fAPAR), while the environmental variables of the study regions are better represented. FLUXCOM also struggles with the only flux tower for West African forests that reported three years of patchy data. Beyond pointing out the sources of data-model discrepancy, this study also highlights that such issues could be fixed if models were better informed with field-measured fAPAR and LUE derived from measured traits, or more generally if better maps of canopy traits are applied[52]. The model simulated GPP could be improved with predicted LUE by optimality theory (Fig. 3), which
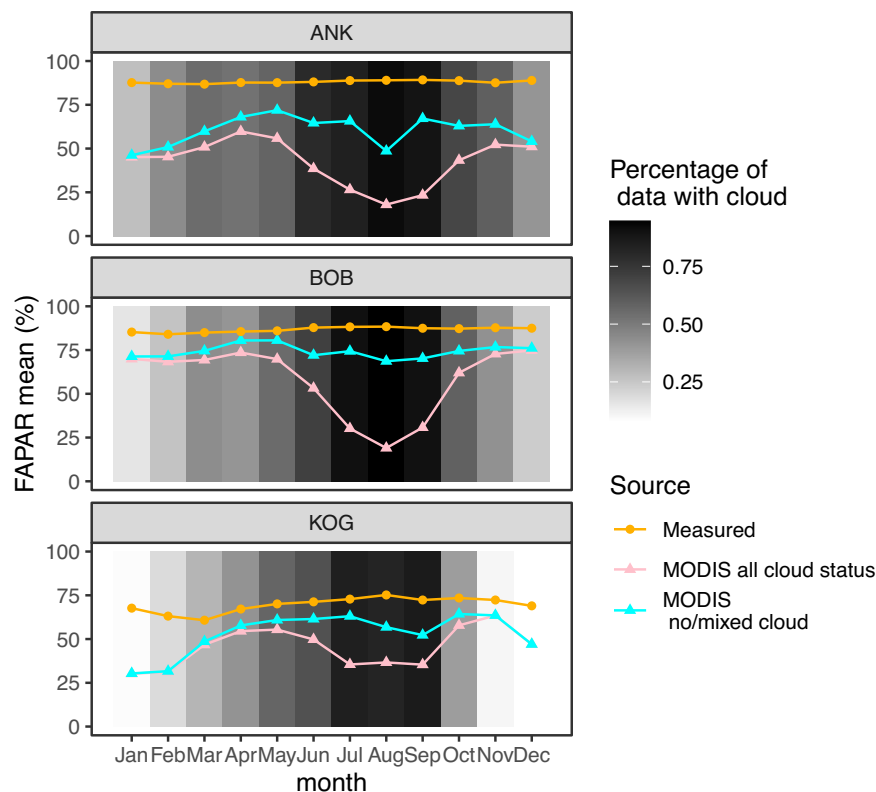
Fig. 4 | The seasonal variation of the fraction of absorbed photosynthetically active radiation (fAPAR). The figure presented data retrieved from MODIS MOD15A2Hv006 (triangle) and in situ measured using hemispherical photography (dot, orange line), at the three study sites (ANK, BOB and KOG), as a monthly mean across 2011 to 2016. The figure shows all MODIS data regardless of the quality flag (Pink line); MODIS fAPAR with cloud contamination control by selecting data marked as 00-no cloud and 10-mixed cloud presents (Blue line). Selecting 00-no cloud only will result in too few data to be presented in the wet season. Grey scale shows the percentage of records contaminated by clouds. Source data are provided as a Source Data file.

highlights the importance of using 'trait-based approaches' parameters instead of PFTs prescription[30,53]. This study thus suggests that providing models with ample field evidence and ensuring strong fidelity to field measurements is critical in improving current carbon cycle simulations[54].

Implication for carbon cycle modelling of West African forests
Researchers studying tropical forest functioning should exercise caution when using satellite-based products subject to cloud contamination, as they are strongly compromised during the rainy season in this region due to high cloudiness (Fig. 1). Moreover, both cloud-free MODIS GPP (Fig. 3) and fAPAR (Fig. 4) are lower than field measurements, suggesting room for improvement beyond cloud contamination. Additionally, West African forests are highly fragmented[55] (Fig. S1), so future studies across scales are suggested to check plant functional types assignment (Fig. S4)[32].

The study has shown that the Pmodel has advanced prediction of c_i/c_a, V_{cmax} and J_{max} – essential parameters for the FvCB model. The data also support the coordination of the Rubisico- and electron transport-limited photosynthetic rates, A_C and A_J, which is one of the optimality principles underlying the Pmodel (Fig. S2). However, the reasons why almost all TRENDY models underestimate West African GPP has not been fully elucidated. The investigation of TRENDY models' LAI and V_{cmax} suggests that these models do not share the same cause of the data-model discrepancy. While much criticism of TRENDY models has centred on likely inaccurate characterisations of plant functional types[30], our study suggests that this is unlikely to be the cause of GPP bias at the study sites. Further, we identified V_{cmax} as a key factor. Numerous models calculate V_{cmax} from leaf nitrogen[30] but

nitrogen-coupled models do not contrast with carbon-only models at the study site (Fig. 2), suggesting that the nitrogen-GPP relationship still has room for improvement in future modelling studies. More transparency in models' input data (e.g. leaf nitrogen content) and parameterization (e.g. V_{cmax}) is necessary for further investigation. Although it is challenging to unbox each model and investigate their V_{cmax}, J_{max} or LUE, our analysis suggests that the underestimation of GPP is associated with photosynthetic capacities, especially at site BOB and KOG (Fig. 2). Allowing V_{cmax}, J_{max} or LUE to acclimate to brighter and drier environments should improve the simulation of GPP at semi-arid forests and savanna[56].

In conclusion, the study not only reveals an underestimation of West African forest productivity but also explores why this underestimation occurs. The unique data-model comparison approach proposed in this study may also have wider potential as it successfully (1) shows consistency between field-measured photosynthetic traits and biometric GPP; (2) identify likely sources of data-model discrepancy by designing multiple experiments with a minimal photosynthesis model (Pmodel). The study also demonstrates that to gain insight, thorough field measurements of forest plots are valuable and the application of simple models (that can be easily understood and tuned) is key for ecological hypothesis testing. As this study is limited in terms of spatial cover, we encourage future research in this region, in particular drier African ecosystems where C4 grasses are abundant with more LUE but less cloudiness (Fig. 1)[57]. We also acknowledge that models are intended for global simulation and thus our site scales study does not serve as models benchmarking but approaches to improve model simulation. Nonetheless, given the broad consistency in our results displayed across the study sites, we expect that models'

carbon cycle simulation in West African region could be substantially improved by simulating a higher GPP across West African tropical forests. It is possible, indeed likely, that such model-data discrepancies are a more general feature of tropical forests. This requires further detailed comparison between biometric field measurements and model predictions, with the approach outlined here offering a valuable approach for such a pantropical analysis.

Methods

Study sites and field measurements

The three study sites span a wet-to-dry rainfall gradient in Ghana from evergreen rainforest at site Ankasa (ANK) with a mean annual precipitation of 2050 mm, to semi-deciduous forest at site Bobiri (BOB) with 1500 mm, and a dry forest and mesic savanna matrix at site Kogyae (KOG) with 1200 mm (Fig. 1). Each study site represents a major forest type of West African forests. See Table S3 for site environmental data.

Biometric GPP was measured in the field using the Global Ecosystem Monitoring protocol. Each component of GPP (e.g. canopy productivity, stem respiration, fine root productivity) is measured separately with common technique. For example, canopy productivity is measured using litterfall trap, and stem respiration is measured by attaching a CO_2 Gas Analyzer (PPsystems EGM4) to stem surface. Biometric GPP was then calculated as a sum of each component[28] and the field practice at our study sites is described in detail in a previous study[22]. Biometric GPP was originally quantified at the plot scale. There are three one-hectare plots at Ankasa, six plots at Bobiri, and five plots at Kogyae. This study was conducted at site rather than plot scale because one-hectare plots at the same site share almost identical climate variables and fall into one model grid cell. Specifically, a grid cell of TRENDY models and FLUXCOM covers approximately 50 × 50 km, or 0.5° × 0.5°. Nonetheless plots at the same site fall into different grid cell of MODIS due to higher resolution (500 ×500 m) (Figure S1). We, therefore, calculated a mean biometric GPP for each site as an average across plots. We did not calculate a standard error. The uncertainty in Biometric GPP (Fig. 2) represents measurements error (mostly systematic error) as calculated by error propagation from each GPP component, instead of spatial variation of GPP. See this study[22] for more data and error propagation associated with the study sites.

The fraction of absorbed photosynthetically active radiation, fAPAR, was obtained using hemispherical photography, taken each month in each plot between 2012 and 2017. Photos were processed using machine learning-based software 'ilastik'[58] for pixel classification and CANEYE[59] for fAPAR calculations (see Supplementary method for details).

In this study, we also need field-measured community-weighted mean light-saturated photosynthetic rate (A_{sat}), light- and CO_2-saturated photosynthetic rate (A_{max}), to calculate a trait-based LUE and trait-based GPP (specific experiment explained in Section 2.4). We derived the maximum carboxylation rate (V_{cmax}) and the maximum electron transport rate (J_{max}) from each field measured A_{sat} and A_{max} at growth temperature[60] using R package 'plantecophys'[61,62]. We used these V_{cmax} and J_{max} in one of the Pmodel experiments for Objective 3. However, the J_{max} limitation equation in 'plantecophys' is different to that in 'rpmodel' (an R package used to conduct Pmodel experiments), so we modified the J_{max} equation in 'plantecophys' (see Supplementary Methods). Measurements of A_{sat} and A_{max} were made every three months from 2014 to 2016 to cover both wet and dry seasons. To measure A_{sat} and A_{max}, we used an open-flow gas exchange system (LI-6400XT, Li-Cor Inc., Lincoln, NE, USA). To ensure a proper representation of the forest stand, we sampled tree species that constituted approximately 80% of the plot basal area. For each species, we selected three mature and canopy emergent trees, and cut one fully sunlit and one shaded branch per tree using a single rope climbing technique. We immediately placed and recut the cut branch under water, and

measured the maximum rate of net CO_2 assimilation at 400 ppm CO_2 (A_{sat}) and 2000 ppm CO_2 (A_{max}) on three leaves per branch. The PPFD was set to 2000 μmol m^{-2} s^{-1} and block temperature was kept constant at 30 °C. Although we measured both shade and sun leaves, we used sun leaves only in this study, consistent with common practice in field studies of photosynthetic traits[63]. Besides, we used above-canopy PPFD as model input and only sun leaves acclimate to this level of PPFD; shade leaves acclimate to darker environments and have consistently lower A_{sat} than sun leaves[64]. The ratio of leaf internal CO_2 to eternal CO_2 (c_i/c_a) was estimated from leaf δ^{13}C measurements. We initially estimated the difference between the leaf stable isotope ratio and the atmospheric stable isotope ratio at that place and time (Δ^{13}C) from δ^{13}C, using the method described by a previous study[65]. Subsequently, we calculated c_i/c_a from Δ^{13}C using equation 11 in a previous study[60]. We compared measured c_i/c_a to Pmodel estimated c_i/c_a (Fig. S3).

MODIS fAPAR and GPP

Using Google Earth Engine, we retrieved MODIS GPP from the collection MOD17A2H and retrieved MODIS fAPAR from MOD15A2H from 2011 to 2016. We extracted GPP and fAPAR of the 14 plots using their coordinates and calculated annual (Fig. 3) and monthly mean values (Fig. 4) per site. To remove cloud-contaminated GPP and fAPAR, we selected only data flagged as 'Significant clouds NOT present (clear)' or 'Mixed cloud present on pixel' provided in band 'faparlaiQC' and 'Psn_QC_500m'. In rain season, almost no data is marked as 'Significant clouds NOT present (clear)' (Figs. 1 and 4). Since MODIS GPP and fAPAR share the same data flag, the above approach ensures that the fAPAR used in Pmodel experiments is identical to that used in MODIS GPP (Fig. 3), ensuring a fair comparison between them. MOD17A3HGF product contains MODIS GPP and NPP after gap-filling MOD17A2H. We used GPP and NPP from MOD17A3HGF to compare with other models and GPP products (Fig. 2, Fig. S1).

FLUXCOM and TRENDY GPP

We chose the 'RS_METEO' version of FLUXCOM because the magnitude of GPP in this version does not involve uncertainty from MODIS fAPAR, which makes the comparison between FLUXCOM and MODIS GPP more independent. Specifically, we used "GPP.RS_METEO.FP-ALL.MLM-ALL.METEO-ERA5.720_360". For TRENDY, we analysed the model outputs in version 9[66] under the S2 protocol, in which climate and CO_2 change while the land cover state is fixed to its pre-industrial state of 1700. Models are classified into nitrogen-carbon coupled models and carbon only models for readers convenience. Note that models still include 'C4 crop land' as one of the plant functional types (PFTs), but the cover of 'C4 crop land' is kept constant. We retrieved total GPP (sum of all PFTs) of the grid cell (variable 'gpp'). We also used variables 'gpppft' and 'landcoverfrac' to calculate a forest-only GPP, which is the GPP that the model would have estimated if the whole grid cell were 'forests', including evergreen, deciduous and any other types of forests. For both TRENDY and FLUXCOM, we extracted the grid cells where the three study sites are located, as an average from 2011 to 2016.

Experiment design for Objective 1

For Objective (1), we compared biometric GPP with values estimated by (a) the TRENDY ensemble of dynamic global vegetation models (DGVMs)[66]; (b) two data-driven products, FLUXCOM[67], and MODIS[68], and (c) an optimality-based model (Pmodel v1.0)[29]. The above choices are widely used models and GPP products that applied distinctly different approaches to calculate GPP[16](Table 2).

The TRENDY v9 S2 ensemble is a collection of 15 DGVMs that calculate functional aspects of vegetation, including fAPAR or leaf area index (LAI), metrics that determine light absorption from environmental variables, without using any remotely sensed data as input.

Most of these models calculate leaf-level photosynthesis via the Farquhar-von Caemmerer-Berry (FvCB) photosynthesis model[69], which requires specification of several photosynthetic traits: V_{cmax}, J_{max}, and parameters of one or other of the semi-empirical schemes that are commonly used to estimate stomatal conductance (g_s)[30]. Leaf-level photosynthesis is scaled up to the canopy, and thus to GPP, by methods that vary in complexity, but all depend on the modelled LAI or fAPAR.

FLUXCOM and MODIS, by contrast, are observation-based GPP products that do not depend on the FvCB model. FLUXCOM GPP is a machine learning application that uses eddy-covariance estimates of GPP from worldwide flux towers as the key input, combined with environmental covariates that include satellite-derived fAPAR, and shortwave radiation (closely related to PPFD). FLUXCOM used MODIS fAPAR, and shortwave radiation prepared by Japan Aerospace eXploration Agency (JAXA) using Terra MODIS data.

MODIS GPP[68] is a light use efficiency model[70,71] that calculates GPP as:

$$GPP = fAPAR * LUE * PPFD \tag{1}$$

where PPFD is sourced from Global Modelling and Assimilation Office (GMAO). The MODIS GPP algorithm calculates LUE as a prescribed (per biome) maximum light use efficiency, multiplied by reduction factors that are defined a priori as biome-specific functions of temperature and vapour pressure deficit.

In this analysis we also employ Pmodel[29]. The Pmodel, uniquely, is a LUE model (also using Eq. 1) but it calculates LUE based on the FvCB model, using optimality principles[72] to determine the spatial and temporal variation in V_{cmax}, J_{max} and the ratio (c_i/c_a) of leaf-internal to ambient CO_2[29]. The c_i/c_a ratio results from the combined effects of photosynthetic rate and stomatal conductance, which are co-regulated by plants. Leaf-level photosynthesis is scaled up to the canopy with the help of the big-leaf approximation[73] and driven by satellite fAPAR data. The Pmodel thus combines the mechanistic basis of photosynthesis as represented in DGVMs with the simplicity of LUE models. The Pmodel also dispenses with the need to consider plant functional type (PFT) or biome distinctions (apart from the difference between C_3 and C_4 plants); the differences in photosynthetic traits among C_3 PFTs are implicitly predicted as a consequence of their habitat preferences alone. The validity of these predictions has been supported by global-scale comparisons[74–76].

Experiment design for Objective 2
Before investigating the cause of the GPP data-model discrepancy (Objective 3), it is necessary to first investigate whether biometric GPP can be reproduced by the FvCB model (within Pmodel) fully informed by field-measured inputs – PPFD, fAPAR and photosynthetic capacities (Objective 2). Here we fed Pmodel with V_{cmax} J_{max} derived from field measured A_{sat} and A_{max} (Supplementary Method). Note that the 'big leaf assumption' is implied for all Pmodel simulation in this study. The above GPP experiment was called 'Pmodel_PfL'. A match would indicate consistency between the field-measured canopy properties and the independent field biometric GPP.

If such a match is found between Pmodel_PfL GPP and biometric GPP, we can further investigate the cause of GPP data-model discrepancy by designing more experiments using Pmodel with different inputs (Table 1 and Fig. 3). The difference between Pmodel_PfL and biometric GPP is labelled as the 'unresolved discrepancy' in Fig. 3.

Experiment design for objective 3
We elucidated four hypotheses for Objective (3) (investigate the cause of data-model discrepancy) (Table 1), which are tested with the following procedure:

To test Hypothesis 1 (incorrect LUE explains the mismatch), we used the Pmodel to predict photosynthetic capacity (Vcmax, Jmax) and LUE using optimality theory based on climate variables alone. The Pmodel then calculates a 'Pmodel_Pf' GPP using the above optimality-based predictions, combined with field-measured PPFD and fAPAR. As the only difference between ' Pmodel_Pf' and 'Pmodel_PfL' GPP is in light use efficiency (derived either from optimality or from in situ measurements), we label the difference in GPP as 'data-model discrepancy resolved by field traits'.

Next, the model calculates a 'Pmodel_P' GPP also using measured PPFD and the above optimality-theory predicted LUE, but with MODIS-derived fAPAR rather than in situ measured fAPAR as input. When 'Pmodel_P' GPP is compared to MODIS GPP, the difference originates from PPFD and LUE (predicted by optimality theory versus from MODIS lookup table). The differences in PPFD among GMAO (used by MODIS[68]), ERA5-Land and field measurements are trivial in explaining GPP bias (Fig. S3) and an intercomparison of climate products is out of the scope of this study. We compared 'Pmodel_P' and 'Pmodel_null' to show that PPFD could not explain the underestimation of GPP at the study sites (Fig. S6). Thus, for simplicity, the difference between 'Pmodel_P' GPP and MODIS GPP is labelled as 'data-model discrepancy resolved by trait optimisation'.

To test Hypothesis 2 (incorrect fAPAR explains the mismatch), we compared 'Pmodel_P' to 'Pmodel_Pf'. The only difference between these two is in fAPAR (in situ measurements versus MODIS). We referred to the difference as 'data-model discrepancy resolved by fAPAR'.

Lastly, 'Pmodel_null' GPP is calculated using satellite PPFD, in which case Pmodel gets no information from field measurements, which is a fair comparison to other models (Fig. 2).

Equivalently to comparing GPP, one could directly compare V_{cmax}, J_{max}, fAPAR and LUE from multiple sources (some presented in Figs. S2 and S3). However, LUE used in MODIS GPP are not publicly available and were difficult to reproduce, so we did not include them. Since the data-model discrepancies in these variables all cascade into GPP, we focus on visualising the discrepancies in GPP in Fig. 3.

Hypothesis 3 (misclassification of land cover) is based on the fact that the GPP estimated from in situ measurements, which is the mean of data from several one-hectare plots, differs in scale from the GPP estimated by the TRENDY models, FLUXCOM and MODIS (as visualised in Fig. S1). West African forests are extremely fragmented, and some forest patches are smaller than the grid cell size of the models. For example, one of the study forests, Bobiri (BOB), is a forest fragment measuring only about 7 × 15 km (Fig. S1), surrounded by cocoa farmland. To illustrate this, we drew a map of MODIS GPP (with a resolution of 500 m) over the Bobiri site, covering an area similar to a single 0.5˚ x 0.5˚ grid cell of the TRENDY models. In some TRENDY models, each grid cell is a composite of multiple PFTs (often represented by different fractions in different models) and those models report GPP per PFT. Thus, we also show the PFT composition in those models of the study sites and compare 'forest-only' GPP to the grid cell average (Fig. S4). The 'forest-only' GPP may contain several types of forests (e.g. deciduous and evergreen) (Table S1). We checked each model documentation to ensure that the 'forest-only' GPP represents the potential GPP of the grid cell if the whole grid cell were forests. LPJ-GUESS was excluded due to its TRENDY documentation, which advises against using landcoverfrac to scale per-PFT data. In LPJ-GUESS, PFTs can overlap due to age-cohorts and understory vegetation, making it impossible to report PFT fluxes in PFT-m² and meet the requirement that the sum equals the land fraction of the grid cell. As a result, PFT fluxes are reported per m²-grid area, not per PFT-m². To calculate gridcell totals, one must sum PFT fluxes across all PFTs (e.g., gpp = sum(gpppft over all PFTs). Any scaling by *landcoverfrac is not meaningful for LPJ-GUESS.

Data-model comparison for climate variables and c_i/c_a

For Objective 3 Hypothesis 4 (i.e. bias in climate variables explain the GPP data-model discrepancy), the following analysis was conducted. We compared temperature and relative humidity from local weather stations to products commonly used by vegetation models and GPP products. TRENDY simulation used CRU-JRA; MODIS GPP used GMAO products, whose latest version is MERRA-2. FLUXCOM reported results forced by different climate variables products and we chose ERA5-Land version. ERA5-Land, which has better resolution than the above, distinguishes climate condition of the study sites well. We chose ERA5-Land to calculate optimality-estimated c_i/c_a via 'rpmodel'. We also derive c_i/c_a from leaf $\delta^{13}C$ measurements, using the method described in ref. [60]. This isotope-derived c_i/c_a was compared to Pmodel-predicted c_i/c_a (Fig. S3). PPFD in this study is calculated from surface incoming shortwave radiation, following[74]. Note that a previous data-model comparison of climate variables for tropical forest sites[77] found ERA-interim outperform Climate Forecast System Reanalysis (CFSR), MERRA2 and the Japanese 55-year Reanalysis (JRA55) for Africa (including our study sites). As we found very small data-model discrepancy in climate variables (Fig. S3) at the study sites, to maintain consistency we used ERA5-Land temperature, vapour pressure and optimality-estimated c_i/c_a for all Pmodel experiments. We avoid using field measured climate variables to inform Pmodel to ensure a fair comparison between Pmodel and other GPP products. The choice of PPFD dataset was specific to the experiment (see simulation schematic in Fig. 3).

Data availability

All data generated in this study have been deposited in the 'figshare' database under accession code [https://doi.org/10.6084/m9.figshare.25431796]. The above includes the environmental data of the study sites. Models and satellite outputs are too large to be uploaded, which will be available upon request. Source data are provided with this paper.

Code availability

All data generated in this study have been deposited in the 'figshare' database under accession code [https://doi.org/10.6084/m9.figshare.25431796]. The above includes the environmental data of the study sites. Models and satellite outputs are too large to be uploaded, which will be available upon request. Source data are provided with this paper.

References

1. Grace, J., Malhi, Y., Meir, P. & Higuchi, N. Productivity of Tropical Rain Forests. in *Terrestrial Global Productivity*. https://doi.org/10.1016/b978-012505290-0/50018-1 (2001).

2. Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).

3. Liu, Y. Y. et al. Recent reversal in loss of global terrestrial biomass. *Nat. Clim. Change* **5**, 470–474 (2015).

4. Ciais, P. et al. Carbon and other biogeochemical cycles. in *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* 465–570 (Cambridge University Press, 2014).

5. Pugh, T. A. M. et al. A large committed long-term sink of carbon due to vegetation dynamics. *Earths Future* **6**, https://doi.org/10.1029/2018EF000935 (2018).

6. Jung, M. et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res. Biogeosciences* **116**, https://doi.org/10.1029/2010JG001566 (2011).

7. Anav, A. et al. Spatiotemporal patterns of terrestrial gross primary production: a review. *Rev. Geophys.* **53**, https://doi.org/10.1002/2015RG000483 (2015).

8. Yang, R. et al. Divergent historical GPP trends among state-of-the-art multi-model simulations and satellite-based products. *Earth Syst. Dyn.* **13**, (2022).

9. Bonan, G. B. et al. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *J. Geophys. Res.* **116**, https://doi.org/10.1029/2010JG001593 (2011).

10. Frankenberg, C. et al. New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophys. Res. Lett.* **38**, https://doi.org/10.1029/2011GL048738 (2011).

11. Chen, M. et al. Regional contribution to variability and trends of global gross primary productivity. *Environ. Res. Lett.* **12**, 105005 (2017).

12. Badgley, G., Anderegg, L. D. L., Berry, J. A. & Field, C. B. Terrestrial gross primary production: Using NIRV to scale from site to globe. *Glob. Change Biol.* **25**, 3731–3740 (2019).

13. Hickler, T. et al. CO2 fertilization in temperate FACE experiments not representative of boreal and tropical forests. *Glob. Change Biol.* **14**, https://doi.org/10.1111/j.1365-2486.2008.01598.x (2008).

14. Wood, T. E., Cavaleri, M. A. & Reed, S. C. Tropical forest carbon balance in a warmer world: A critical review spanning microbial- to ecosystem-scale processes. *Biol. Rev.* **87**, 912–927 (2012).

15. Babst, F. et al. Modeling ambitions outpace observations of forest carbon allocation. *Trends Plant Sci.* **26**, 210–219 (2021).

16. Zhang, Y. & Ye, A. Would the obtainable gross primary productivity (GPP) products stand up? A critical assessment of 45 global GPP products. *Sci. Total Environ.* **783**, 146965 (2021).

17. Joiner, J. et al. Estimation of terrestrial global gross primary production (GPP) with satellite data-driven models and eddy covariance flux data. *Remote Sens* **10**, 1346 (2018).

18. Tian, Z. et al. Fusion of multiple models for improving gross primary production estimation with eddy covariance data based on machine learning. *J. Geophys. Res. Biogeosciences*. e2022JG007122 https://doi.org/10.1029/2022JG007122 (2023).

19. Wang, L. et al. Evaluation of the latest MODIS GPP products across multiple biomes using global eddy covariance flux data. *Remote Sens.* **9**, https://doi.org/10.3390/rs9050418 (2017).

20. Seiler, C. et al. Are terrestrial biosphere models fit for simulating the global land carbon sink? *J. Adv. Model. Earth Syst.* **14**, e2021MS002946 (2022).

21. Ardö, J. Comparison between remote sensing and a dynamic vegetation model for estimating terrestrial primary production of Africa. *Carbon Balance Manag* **10**, 8 (2015).

22. Zhang-Zheng, H. et al. Contrasting carbon cycle along tropical forest aridity gradients in West Africa and Amazonia. *Nat. Commun.* **15**, 3158 (2024).

23. Prentice, I. C. & Cowling, S. A. Dynamic global vegetation models. in *Encyclopedia of biodiversity* (ed. Levin, S. A.) 670–689 (Elsevier, Amsterdam). https://doi.org/10.1016/B978-0-12-384719-5.00412-3, (2013).

24. Franklin, O. et al. Organizing principles for vegetation dynamics. *Nat. Plants* **6**, 444–453 (2020).

25. Moore, S. et al. Forest biomass, productivity and carbon cycling along a rainfall gradient in West Africa. *Glob. Change Biol.* **24**, e496–e510 (2018).

26. Gvozdevaite, A. et al. Leaf-level photosynthetic capacity dynamics in relation to soil and foliar nutrients along forest–savanna boundaries in Ghana and Brazil. *Tree Physiol.* **38**, 1912–1925 (2018).

27. Oliveras, I. et al. The influence of taxonomy and environment on leaf trait variation along tropical abiotic gradients. *Front. Glob. Change* **3**, 18 (2020).

28. Malhi, Y. et al. The Global Ecosystems Monitoring network: Monitoring ecosystem productivity and carbon cycling across the tropics. *Biol. Conserv.* **253**, 108889 (2021).

29. Stocker, B. D. et al. P-model v1.0: An optimality-based light use efficiency model for simulating ecosystem gross primary production. *Geosci. Model Dev.* **13**, 1545–1581 (2020).

30. Rogers, A. et al. A roadmap for improving the representation of photosynthesis in Earth system models. *N. Phytol.* **213**, 22–42 (2017).

31. Luo, X. et al. Comparison of big-leaf, two-big-leaf, and two-leaf upscaling schemes for evapotranspiration estimation using coupled carbon-water modeling. *J. Geophys. Res. Biogeosciences* **123**, 207–225 (2018).

32. Morel, A. C. et al. Carbon dynamics, net primary productivity and human-appropriated net primary productivity across a forest–cocoa farm landscape in West Africa. *Glob. Change Biol.* **25**, 2661–2677 (2019).

33. Domingues, T. F. et al. Co-limitation of photosynthetic capacity by nitrogen and phosphorus in West Africa woodlands. *Plant Cell Environ.* **33**, 959–980 (2010).

34. Tomlinson, K. W. et al. Leaf adaptations of evergreen and deciduous trees of semi-arid and humid savannas on three continents. *J. Ecol.* **101**, 430–440 (2013).

35. Belelli Marchesini, L. et al. Ankasa Flux Tower: A New Research Facility for the Study of the Carbon Cycle in a Primary Tropical Forest in Africa. **2011**, 13–23 (2011).

36. Hayek, M. N. et al. A novel correction for biases in forest eddy covariance carbon balance. *Agric. Meteorol.* **250–251**, 90–101 (2018).

37. Billesbach, D. P. et al. Effects of the Gill-Solent WindMaster-Pro "w-boost" firmware bug on eddy covariance fluxes and some simple recovery strategies. *Agric. Meteorol.* **265**, 145–151 (2019).

38. de Araújo, A. C. et al. The spatial variability of CO2 storage and the interpretation of eddy covariance fluxes in central Amazonia. *Agric. Meteorol.* **150**, 226–237 (2010).

39. Fu, Z. et al. The surface-atmosphere exchange of carbon dioxide in tropical rainforests: Sensitivity to environmental drivers and flux measurement methodology. *Agric. Meteorol.* **263**, 292–307 (2018).

40. Xu, L. et al. Satellite observation of tropical forest seasonality: spatial patterns of carbon exchange in Amazonia. *Environ. Res. Lett.* **10**, 084005 (2015).

41. Rödig, E. et al. The importance of forest structure for carbon fluxes of the Amazon rainforest. *Environ. Res. Lett.* **13**, 054013 (2018).

42. Zhao, M., Heinsch, F. A., Nemani, R. R. & Running, S. W. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.* **95**, 164–176 (2005).

43. Fensholt, R. & Proud, S. R. Evaluation of Earth Observation based global long term vegetation trends - Comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.* **119**, 131–147 (2012).

44. Fuster, B. et al. Quality Assessment of PROBA-V LAI, fAPAR and fCOVER Collection 300 m Products of Copernicus Global Land Service. *Remote Sens* **12**, 1017 (2020).

45. Sudmanns, M., Tiede, D., Augustin, H. & Lang, S. Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass. *Int. J. Digit. Earth* **13**, 768–784 (2020).

46. Chen, J. M. et al. Global datasets of leaf photosynthetic capacity for ecological and earth system research. *Earth Syst. Sci. Data* **14**, 4077–4093 (2022).

47. Dong, N. et al. Rising CO2 and warming reduce global canopy demand for nitrogen. *N. Phytol.* **235**, 1692–1700 (2022).

48. Sibret, T. et al. High photosynthetic capacity of Sahelian C3 and C4 plants. *Photosynth. Res.* **147**, 161–175 (2021).

49. Lu, X., Croft, H., Chen, J. M., Luo, Y. & Ju, W. Estimating photosynthetic capacity from optimized Rubisco–chlorophyll relationships among vegetation types and under global change. *Environ. Res. Lett.* **17**, 014028 (2022).

50. Rogers, A. The use and misuse of Vc,max in Earth System Models. *Photosynth. Res.* **119**, 15–29 (2014).

51. Kattge, J., Knorr, W., Raddatz, T. & Wirth, C. Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Glob. Change Biol.* **15**, 976–991 (2009).

52. Aguirre-Gutiérrez, J. et al. Pantropical modelling of canopy functional traits using Sentinel-2 remote sensing data. *Remote Sens. Environ.* **252**, 112122 (2021).

53. Wullschleger, S. D. et al. Plant functional types in Earth system models: past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. *Ann. Bot.* **114**, 1–16 (2014).

54. Prentice, I. C., Liang, X., Medlyn, B. E. & Wang, Y. P. Reliable, robust and realistic: The three R's of next-generation land-surface modelling. *Atmos. Chem. Phys.* **15**, 5987–6005 (2015).

55. Buchanan, G. M., Field, R. H., Bradbury, R. B., Luraschi, B. & Vickery, J. A. The impact of tree loss on carbon management in West Africa. *Carbon Manag* **12**, 623–633 (2021).

56. Zhang-Zheng, H. et al. Photosynthetic and water transport strategies of plants along a tropical forest aridity gradient: a test of optimality theory. 2023.01.10.523419 Preprint at https://doi.org/10.1101/2023.01.10.523419 (2023).

57. Chen, Y., Feng, X., Fu, B., Wu, X. & Gao, Z. Improved Global Maps of the Optimum Growth Temperature, Maximum Light Use Efficiency, and Gross Primary Production for Vegetation. *J. Geophys. Res. Biogeosciences* **126**, e2020JG005651 (2021).

58. Berg, S. et al. Ilastik: interactive machine learning for (bio) image analysis. *Nat. Methods* **16**, 1226–1232 (2019).

59. Weiss, M. & Baret, F. CAN_EYE V6. 4.91 user manual. 1–56 (2017).

60. Peng, Y., Bloomfield, K. J. & Prentice, I. C. A theory of plant function helps to explain leaf-trait and productivity responses to elevation. *N. Phytol.* **226**, 1274–1284 (2020).

61. Duursma, R. A. Plantecophys-an R package for analysing and modelling leaf gas exchange data. *PloS One* **10**, e0143346 (2015).

62. De Kauwe, M. G. et al. A test of the 'one-point method'for estimating maximum carboxylation capacity from field-measured, light-saturated photosynthesis. *N. Phytol.* **210**, 1130–1144 (2016).

63. Cornelissen, J. H. C. et al. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Aust. J. Bot.* **51**, 335–380 (2003).

64. Keenan, T. F. & Niinemets, Ü. Global leaf trait estimates biased due to plasticity in the shade. *Nat. Plants* **3**, 1–6 (2016).

65. Cornwell, W. wcornwell/leaf13C v0.1.0 (v0.1.0). *Zenodo* https://doi.org/10.5281/zenodo.569501 (2017).

66. Friedlingstein, P. et al. Global Carbon Budget. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).

67. Jung, M. et al. The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Sci. Data* **6**, 74 (2019).

68. Running, S., Mu, Q. & Zhao, M. MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC https://doi.org/10.5067/MODIS/MOD17A2H.006 (2015).

69. Farquhar, G. D., von Caemmerer, S. & Berry, J. A. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta* **149**, 78–90 (1980).

70. Hilker, T., Coops, N. C., Wulder, M. A., Black, T. A. & Guy, R. D. The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements. *Sci. Total Environ.* **404**, 411–423 (2008).

71. Sun, Y. et al. Impact of mesophyll diffusion on estimated global land CO2 fertilization. *Proc. Natl Acad. Sci.* **111**, 15774–15779 (2014).

72. Harrison, S. P. et al. Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *N. Phytol.* **231**, 2125–2141 (2021).

73. Friend, A. Modelling canopy CO2 fluxes: are 'big-leaf'simplifications justified? *Glob. Ecol. Biogeogr.* **10**, 603–619 (2001).

74. Wang, H. et al. Towards a universal model for carbon dioxide uptake by plants. *Nat. Plants 2017 39* **3**, 734–741 (2017).
75. Smith, N. G. et al. Global photosynthetic capacity is optimized to the environment. https://doi.org/10.1111/ele.13210 (2019).
76. Peng, Y., Bloomfield, K. J., Cernusak, L. A., Domingues, T. F. & Colin Prentice, I. Global climate and nutrient controls of photosynthetic capacity. *Commun. Biol. 2021 41* **4**, 1–9 (2021).
77. Burton, C., Rifai, S. & Malhi, Y. Inter-comparison and assessment of gridded climate products over tropical forests during the 2015/2016 El Niño. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170406 (2018).

Author contributions
H.Z., I.C.P., Y.M., conceived the study. S.A.B., A.D., A.G., S.M., and I.O. collected field data. R.D., E.T., B.S., and H.Z. did the modelling work. H.Z., X.D. carried out data analysis. J.A. revised the manuscript. H.Z. wrote the paper with inputs and revisions from all co-authors.

Competing interests
The authors declare no competing interests.

Inclusion & Ethics statement
Local researchers were involved throughout the research process, with clear roles and responsibilities agreed upon, particularly with contributions from the Forestry Research Institute of Ghana. The research was locally relevant and determined in collaboration with local partners. Capacity-building workshops were organised during the research process. Local and regional research relevant to the study is cited, as evidenced by the references to West Africa specific studies. Risk management plans were made for the fieldwork following the universities' regulations.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53949-0.

Correspondence and requests for materials should be addressed to Huanyuan Zhang-Zheng, I. Colin Prentice or Yadvinder Malhi.

Peer review information *Nature Communications* thanks Christian Seiler and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.