This conceptual model implies that GPP and stand age may intrinsically determine carbon allocation of GPP to NEP in global forests. ❒

References
1. Fernández-Martínez, M. *et al. Nature Clim. Change* **4,** 471–476 (2014).
2. Goulden, M. L. *et al. Global Change Biol.* **17,** 855–871 (2011).
3. Odum, E. P. *Science* **164,** 262–270 (1969).
4. Ryan, M. G. & Yoder, B. J. *Bioscience* **47,** 235–242 (1997).
5. Johnson, D. W. *Ecology* **87,** 64–75 (2006).
6. Luyssaert, S. *et al. Nature* **455,** 213–215 (2008).

Enzai Du[1,2]
[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, and College of Resources Science & Technology, Beijing Normal University, Xinjiekouwai Street 19, Beijing 100875, China, [2]Environmental Systems Analysis Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands.
*e-mail: enzaidu@bnu.edu.cn

CORRESPONDENCE:

Data quality and the role of nutrients in forest carbon-use efficiency

To the Editor — Predictions of future forest carbon storage are uncertain because of restricted knowledge about drivers of forest carbon cycling. Fernández-Martínez *et al.*[1] state that nutrient availability is the key regulator of global forest carbon balance. This conclusion was drawn from carbon balances of 92 forests mainly derived from eddy covariance data. The key variable was ecosystem carbon-use efficiency (CUEe), defined as ratio of net ecosystem carbon uptake (NEP) to gross primary production (GPP). In their study, comparing ecosystems with high, medium and low nutrient availability resulted in a fivefold higher average CUEe for nutrient-rich forests. Our re-analysis shows, however, that the underlying data set contained flawed data, and the study ignored factors such as site history and topographical site characteristics that influence the quality of eddy covariance data. Including these factors as a quality control results in a data set that does not show any significant influence of nutrient availability on CUEe.

Our re-analysis focused on three aspects of the data (for details see Supplementary Information).

Is the quality of the data of the same high standard for all sites? For this purpose, it is important to understand the provenance of the data used in the study. The final data set (FMD, one average per site) is an extract from a global forest database[2] (SLD, several annual values per site) built on literature data and extracts from databases of international networks (EFDC, half-hourly values). All steps in this data chain were partly re-checked. Unreliable data were found in all three data sets, and 11 sites had to be removed.

In a second step, data from very young forests were re-analysed because the authors considered age but not previous history.

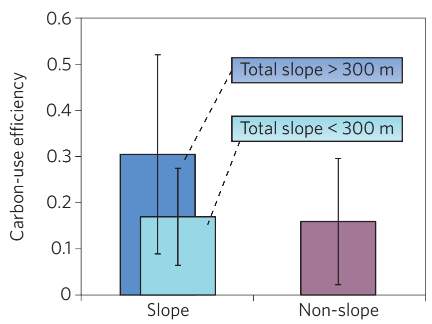

Figure 1 | Average CUEe for sites in complex terrain compared with sites in flat terrain. Sites in complex terrain were split into sites with differences in altitude >300 m and <300 m.

However, history strongly influences ecosystem carbon balances, mainly via soil carbon stocks, whereas nutrient availability plays only a minor role[3]. All sites aged younger than 15 years old were binned into 'afforestation', 'disturbance' or 'unknown'. Each group showed high correlation between CUEe and age, but the group 'afforestation' showed an initially high CUEe that decreased thereafter, whereas the forests with 'disturbance' had a highly negative CUEe at the beginning that increased thereafter. Around an age of 15 years the difference between the two groups vanished. This is in accordance with recent observations on land-use changes[4]. Twenty-three sites younger than 15 years were excluded from the re-analysis because the authors did not consider the site history of young forests in their statistics.

Terrain features were analysed for the remaining 95 sites. Complex terrain has been a focus of research for some time, but even substantial efforts to understand its influence on flux measurements[5] did not lead to a clear description of the phenomenon[6]. Therefore, the basic hypothesis of most of the FLUXNET community became that complex terrain probably causes a random but not a systematic error. Across a large number of sites, this would balance out and could therefore be ignored. This hypothesis was tested in this study: if terrain causes random error, CUEe should not correlate with any parameter describing the terrain around flux towers, and average CUEe from towers in complex terrain should not differ from that from towers in flat terrain. Forty-four sites were identified as being located in complex terrain. For these sites, CUEe was correlated to the total difference in altitude of the terrain (TDA). A positive correlation between TDA and CUEe was found, and average CUEe was significantly higher for sites with more than 300 m TDA (Fig. 1). Therefore, this difference in altitude was taken as threshold for a severe influence of advection, and 13 sites were excluded. Supplementary Table S4 details the sites removed during the different data quality checks and correction steps.

The remaining sites ($n = 82$) were re-analysed. Plotting Re against GPP gave a surprisingly high correlation throughout all sites ($r^2 = 0.899$, Fig. 2a). Only a small but not significant influence of nutrient availability on CUEe was detected (Fig. 2b).

The study by Fernández-Martínez *et al.*[1] shows clearly the necessity of improvements in integration studies related to eddy-covariance-derived data.

Data curation has to be improved. Some data were too low-quality for the purpose of this study. This shows the importance of high standards of data quality and is a big challenge for further integration (for example within FLUXNET). Infrastructures such as ICOS, Asiaflux, Ameriflux, Chinaflux, Ozflux or NEON need to ensure the highest standards of data quality and provide centralized, transparent and well-documented procedures for data processing, provenance and metadata. Data contributors from outside these infrastructures should receive clear guidelines, and their data should only be accepted when they can prove that they have followed the guidelines.

The community must accept that complex terrain cannot be ignored in integration studies. The basic hypothesis that different influences of terrain balance each other out in big data sets has been falsified here, at least for towers located in terrain with TDA higher than 300 m. Sites in terrain with lower TDA have also been shown to be affected by terrain[7], but with the relatively simple approach used here this cannot be proved statistically. More effort must be invested here, and until the problem of complex terrain has been solved, infrastructures and integration networks should clearly communicate possible terrain influences to data users.

Finally, it is important to consider all important factors and additional information when deriving general ecological hypotheses. In the present case, fertilization experiments (for example on clearcuts) could support the various ideas.

Overall, the re-analysis shows that the ecological conclusions drawn by Fernández-Martínez *et al.*[1] are not justified. Nevertheless, the re-analysis also shows that the eddy covariance method as such, although not applicable in all terrains, allows important insights into the ecology of forest ecosystems. The most important result is the strong correlation between GPP and ecosystem respiration. A CUEe between 0 and 0.3 with an average around 0.15 may be a reasonable result from this data set for modellers. An ensemble of other factors is likely to influence CUEe within this range. Nutrient availability is certainly one of them[8], but not as unequivocally as claimed. ❐

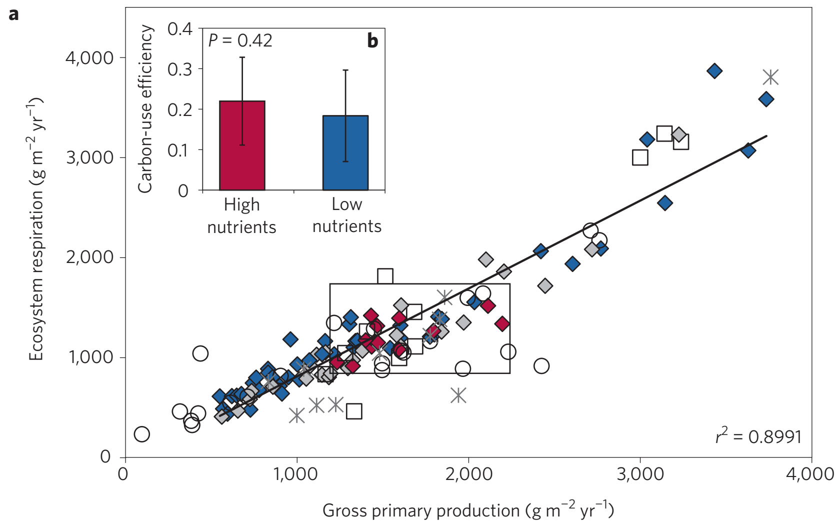

Figure 2 | Ecosystem respiration (Re) plotted against GPP for the remaining 82 sites. **a**, Red: sites with high nutrient availability. Blue: sites with low nutrient availability. Grey: sites with medium nutrient availability. Open squares: sites removed owing to bad data quality and unclosed carbon balance that could not be fixed. Open circles: removed sites younger than 15 years. Grey stars: removed sites with complex terrain. **b**, Average CUEe for sites with low and high nutrient availability with a GPP between 1,200 and 2,300 gC m^{-2} yr^{-1}.

References

1. Fernández-Martínez, M. *et al. Nature Clim. Change* **4,** 471–476 (2014).
2. Luyssaert, S. *et al. Glob. Change Biol.* **13,** 2509–2537 (2007).
3. Grant, R. F. *et al. Tellus Ser. B* **62,** 475–496 (2010).
4. Poeplau, C. & Don, A. *Geoderma* **192,** 189–201 (2013).
5. Belcher, S. E., Harman, I. N. & Finnigan, J. J. *Annu. Rev. Fluid Mech.* **44,** 479–504 (2012).
6. Feigenwinter, C. *et al. Agric. Forest Meteorol.* **148,** 12–24 (2008).
7. Kutsch, W. L. *et al. Ecol. Applic.* **18,** 1391–1405 (2008).
8. Vicca, S. *et al. Ecol. Lett.* **15,** 520–526 (2012).

Acknowledgements

We thank the numerous site PIs and their co-workers, as well as their funding institutions. M. Fernández-Martínez supported this study by providing his data set. We thank T. Vesala, J. Finnigan and J. Grace for comments on an earlier draft of this manuscript. W.K. was supported through grant 281250 (ICOS-ERIC) by the Academy of Finland. P.K. was supported through the following grants by the Academy of Finland: 272041 (Center of Excellence), 271878 (ICOS) and 281255 (ICOS-Finland).

Additional information

Supplementary information is available in the online version of the paper.

Werner L. Kutsch[1]* and Pasi Kolari[2]*

[1]Integrated Carbon Observation System (ICOS RI), Interim Headoffice, Erik Palménin aukio 1, PO Box 48, FI-00014 University of Helsinki, Finland; Department of Physics, University of Helsinki, Erik Palménin aukio 1, PO Box 48, FI-00014 University of Helsinki, Finland. *e-mail: werner.kutsch@helsinki.fi; pasi.kolari@helsinki.fi

Reply to 'Uncertain effects of nutrient availability on global forest carbon balance' and 'Data quality and the role of nutrients in forest carbon-use efficiency'

Fernández-Martínez *et al.* reply — Du suggested in his Correspondence[1] that our analysis[2] was flawed for several reasons and offered a new hypothesis. Our analyses and conclusions were not based on the simple regression presented in Fig. 1 of our paper[2]. The figure was merely meant for visualization purposes, showing the data and the differences between fertile and infertile sites. We relied instead on generalized linear models (GLMs; see Supplementary Information in ref. 2). Our study showed that NEP was affected not only by fertility and GPP, but also by stand age, mean annual temperature, water deficit and management (Table 1 of ref. 2). Conclusions therefore cannot be based on linear regressions restricted to a partial set of predictor variables. Stand age in our models in fact interacted with GPP and therefore presented a nonlinear relationship with NEP, precisely as Du suggests in his conceptual model. The Correspondence further claims that three young forests with the highest carbon-use efficiency (CUEe) confounded our analysis. This claim is incorrect. Our analyses were supported by leverage tests[3], which showed that these sites did not affect our results. Nonetheless, as shown in the Supplementary Information of ref. 2, we repeated all analyses using only data from the eddy covariance towers (excluding these three sites with the highest CUEe), and yet the patterns remained unchanged. Similarly, the comment suggested the use of different GPP ranges, but all analyses in the original paper also excluded all high-GPP