

Streamflow response to forest management

Kirchner, James W.; Berghuijs, Wouter R.; Allen, Scott T.; Hrachowitz, Markus; Hut, Rolf; Rizzo, D.M.

DOI

[10.1038/s41586-020-1940-6](https://doi.org/10.1038/s41586-020-1940-6)

Publication date

2020

Document Version

Final published version

Published in

Nature

Citation (APA)

Kirchner, J. W., Berghuijs, W. R., Allen, S. T., Hrachowitz, M., Hut, R., & Rizzo, D. M. (2020). Streamflow response to forest management. *Nature*, 578(7794), E12-E15. <https://doi.org/10.1038/s41586-020-1940-6>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Streamflow response to forest management

<https://doi.org/10.1038/s41586-020-1940-6>

Received: 24 July 2019

Accepted: 2 December 2019

Published online: 12 February 2020

James W. Kirchner^{1,2*}, Wouter R. Berghuijs¹, Scott T. Allen^{1,3}, Markus Hrachowitz⁴, Rolf Hut⁴ & Donna M. Rizzo⁵

Arising from: Evaristo, J. & McDonnell, J. J. *Nature* <https://doi.org/10.1038/s41586-019-1306-0> (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019); Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019); Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).

Forests play a key part in the water cycle, so both planting and removing forests can affect streamflow. In a recent Article¹, Evaristo and McDonnell used a gradient-boosted-tree model to conclude that streamflow response to forest removal is predominantly controlled by the potential water storage in the landscape, and that removing the world's forests would contribute an additional 34,098 km³ yr⁻¹ to streamflow worldwide, nearly doubling global river flow. Here we report several problems with Evaristo and McDonnell's¹ database, their model, and the extrapolation of their results to the continental and global scale. The main results of the paper¹ remain unsubstantiated, because they rely on a database with multiple errors and a model that fails validation tests.

Database problems

We spot-checked the database underlying Evaristo and McDonnell's analysis¹ by comparing individual entries to the original cited references. Roughly half of these spot checks revealed substantial errors in the calculated changes in water yields, or errors in the classification of individual studies as forest planting versus forest removal experiments. Here we describe four examples. (1) The Valtorto catchment in Portugal is classified as a forest clearing experiment¹ although the catchment was never forested, but rather covered by 50-cm-tall heath². The reported post-clearing streamflow increase of 363.6% (ref.¹) is also inconsistent with table 3 of ref.², which reports that average streamflow increased by 150%, from 1.0 m³ per day to 2.5 m³ per day. (2) The database reports that forest clearing at the Lemon catchment in Australia increased streamflow by 631.8% (ref.¹), but from table 1 of ref.³, we calculate that the average pre- and post-clearing streamflows were 18.0 mm yr⁻¹ and 27.9 mm yr⁻¹ respectively, implying that streamflow increased by only 55%. (3) Brigalow catchments C2 and C3, which each appear twice in the database, are classified as forest planting experiments¹ although neither was planted with forest: C2 was planted with sorghum and wheat and C3 was planted with buffel grass for pasture^{4,5}. (4) Several forest conversion experiments, in which forests were cleared and replanted with other vegetation (for example, references 74, 114, 130 and 163 in ref.¹), are reported in the database as showing, counterintuitively, large streamflow increases caused by forest planting¹. However, the reported changes in streamflow were calculated relative to intact forest control plots, not cleared land, so they mostly reflect the effects of clearing the existing forest rather than the effects of planting. We suspect that this misattribution of forest clearing effects to forest planting may underlie the paper's surprising finding (see Fig. 2 of ref.¹ and associated discussion) that forest planting appears to increase streamflow by 100% or more at many sites, with the largest increases at sites with

the highest evapotranspiration rates, a pattern that would normally arise from forest clearing instead.

Model overfitting and validation failure

Gradient-boosted regression trees are data-hungry, and although Evaristo and McDonnell¹ compiled every paired watershed study that they could find, the resulting databases of 161 forest clearing experiments and 90 forest planting experiments are much too small to estimate their seven-variable model reliably. We checked the model codes that Evaristo and McDonnell provided with their paper (see the code availability statement of ref.¹) and found that the boosted tree algorithm fits 200 free parameters (not counting the dozens of additional free parameters that define the tree's branch points), suggesting substantial overfitting. To test how this overfitting might affect the model's predictions, we split the forest removal and planting databases into training sets (80% of the data) and test sets (the remaining 20% of the data). To balance the distributions of the variables between the training and test sets, we used stratified random sampling; we also used un-stratified random sampling as a more stringent test. We then re-ran the boosted-tree analysis, using the same data, the same platform (JMP, the SAS Institute), and the same algorithm options that Evaristo and McDonnell¹ used, for 300 of these random splits of the data, both with and without 'early stopping' (in which the fitting algorithm stops whenever the next layer would reduce the R^2).

The results in Fig. 1 show that the model fails these validation tests. If the model were not overfitted, the fits to the test data (as measured by the test R^2 on the vertical axis) would be similar to the fits to the training data (as measured by the training R^2 on the horizontal axis), and the dots would lie close to the 1:1 line. Instead, many of the dots lie far below the 1:1 line, and many test R^2 values even lie below zero, indicating model predictions that are worse than random guessing. Figure 1 thus shows that the model is overfitted and makes unreliable predictions (because it is too flexible, and thus has been 'fitted to the noise' in the training data). This result holds whether one uses 'early stopping' or not, and both stratified and un-stratified validation tests yield broadly similar results.

Although individual randomizations can yield test R^2 values that are similar to the training R^2 (or even higher), one should not draw conclusions from such anomalies. Model performance is better reflected in the medians of the training and test R^2 values across many randomization trials (Table 1). Table 1 confirms quantitatively what Fig. 1 shows visually: in each case, the median test R^2 is much smaller than the median training R^2 , and many test R^2 values are below zero.

¹Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland. ²Swiss Federal Research Institute WSL, Birmensdorf, Switzerland. ³Department of Geology and Geophysics, University of Utah, Salt Lake City, UT, USA. ⁴Department of Civil Engineering, Delft University of Technology, Delft, The Netherlands. ⁵Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT, USA. *e-mail: kirchner@ethz.ch

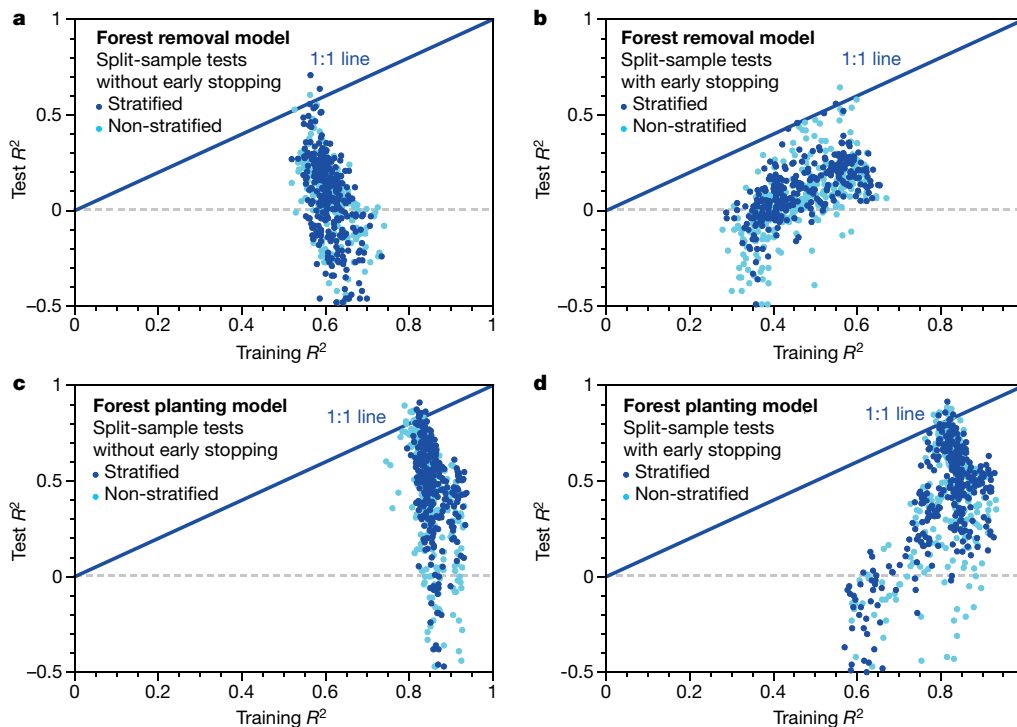


Fig. 1 | Split-sample validation tests of gradient-boosted-tree model fitted to forest clearing and planting data. **a, b,** Model fitted to forest clearing data with and without early stopping; **c, d,** model fitted to forest planting data with and without early stopping. The source data were randomly split into 300 training and test sets in 80/20 ratios, as described in the text. If the model were

not overfitted, the R^2 statistics obtained from the training and test sets would be similar to one another, and thus the dots would lie close to the 1:1 lines. Instead, the test R^2 statistics are generally much smaller than the training R^2 values. Points with test R^2 values less than -0.5 , which indicate that model predictions were much worse than random guessing, are not shown.

All of the paper's¹ main results are based on the boosted-tree model, so the validation failure documented here invalidates the paper's conclusions. The other machine learning methods in the paper have similar validation issues, but we will not explore them in detail because the paper's conclusions do not depend on them.

Exaggerated importance of potential storage

The finding¹ that streamflow response to forest removal was primarily controlled, not by climate, but by total potential water storage in the landscape, was puzzling to us for two reasons. First, it was difficult to imagine how total storage, much of which may lie below the rooting

zone of trees, could be the major control on the hydrological effects of tree removal. Second, given that forest planting and forest removal both alter the same variable (forest cover), but in opposite directions, it was hard to reconcile the paper's two main findings¹: that potential storage is the dominant control on streamflow response to forest clearing (but not planting), and that actual evapotranspiration (AET) is the dominant control on streamflow response to forest planting (but not clearing).

Closer examination reveals that the apparent importance of potential storage relies on one extreme data point (the Lemon catchment, Australia), which has a potential storage of 15 m, more than twice the next-highest value in the dataset. If we remove this one data point, potential storage disappears as the most important factor (Table 2), and is replaced by potential evapotranspiration (PET). This one data point is so influential because Evaristo and McDonnell's analysis¹ uses an 'independent uniform' variable importance profiler. This profiler is intended for use where the likely values of each variable will be uniformly distributed over the range of the data⁶, which is inconsistent with the strongly skewed distributions of potential storage in Evaristo and McDonnell's paired watershed dataset (Fig. 2a) and in their global catchment database (Fig. 2b). Potential storages exceeding 7.5 m comprise only 0.6% of Evaristo and McDonnell's paired watershed dataset (light blue bars, Fig. 2a) and 6% of their global catchment database (light blue bars, Fig. 2b), but 50% of the distribution used to calculate the influence of potential storage, exaggerating its importance.

Although Evaristo and McDonnell fully documented their choice of this "independent uniform" profiler¹, other choices, more consistent with the available data, lead to a different conclusion. For example, if we instead use a profiling method that takes into account the actual distributions of all of the variables ("independent resampled" profiling), PET becomes the most important variable, and potential storage drops to fourth place (Table 2). And if the profiling method also takes account of the correlations among the variables, in addition to their actual

Table 1 | Summary of split-sample validation test results

Model and split-sample test performed (80/20 split in all cases)	Median training R^2	Median test R^2	Fraction of test $R^2 < 0$
Forest removal model			
Stratified, with early stopping	0.449	0.108	31%
Stratified, without early stopping	0.605	0.096	36%
Unstratified, with early stopping	0.458	0.053	34%
Unstratified, without early stopping	0.608	0.057	40%
Forest planting model			
Stratified, with early stopping	0.827	0.455	13%
Stratified, without early stopping	0.852	0.486	10%
Unstratified, with early stopping	0.826	0.475	16%
Unstratified, without early stopping	0.844	0.474	17%

Test results are shown for the boosted-tree model fitted to forest removal and forest planting data. 'Fraction of test $R^2 < 0$ ' indicates the percentage of tests in which model predictions were worse than random guessing.

Table 2 | Relative variable importance using different profilers

Profiling method and treatment of Lemon catchment	Potential evapotranspiration	Runoff coefficient	Drainage area	Potential storage	Actual evapotranspiration	Root zone storage	Permeability
Independent uniform							
Lemon included	0.317 (2)	0.098 (3)	0.036 (5)	0.508 (1)	0.041 (4)	0.007 (6)	0.000 (7)
Lemon omitted	0.500 (1)	0.056 (4)	0.031 (5)	0.299 (2)	0.179 (3)	0.001 (6)	0.001 (6)
Independent resampled							
Lemon included	0.642 (1)	0.114 (3)	0.165 (2)	0.094 (4)	0.030 (5)	0.005 (6)	0.000 (7)
Lemon omitted	0.710 (1)	0.077 (4)	0.134 (2)	0.091 (3)	0.050 (5)	0.001 (6)	0.003 (7)
Dependent resampled							
Lemon included	0.440 (1)	0.189 (2)	0.171 (3)	0.137 (5)	0.109 (6)	0.155 (4)	0.095 (7)
Lemon omitted	0.433 (1)	0.180 (2)	0.174 (3)	0.129 (5)	0.102 (6)	0.161 (4)	0.098 (7)

Relative importance scores for each of the seven variables in Evaristo and McDonnell's forest removal model¹ are shown for three different profiling methods, and for including and excluding the Lemon catchment (see text). Ranks are shown in parentheses. The most important variable in each case is highlighted in bold.

distributions (“dependent resampled” profiling), the most important variable is again PET, and potential storage drops to fifth place out of seven variables (regardless of whether we include or exclude the Lemon catchment; see Table 2).

Exaggerated global streamflow implications

To estimate the potential impact of forest clearing on global streamflow (table 1 of ref. ¹), Evaristo and McDonnell first applied their boosted-tree model to a database of 442,319 catchments for which the required seven input variables are available (whether or not they are actually forested). Evaristo and McDonnell then multiplied the median of the modelled percentage change in streamflow for each continent's catchments by the average continental river flow (see Table 3). Because less than 30% of Earth's land area is forested⁷, however, the potential percentage increase in streamflow from forest clearing should not be applied to the entire continental runoff; that is, one cannot clear forests from the 70% of Earth's land surface where no forests exist. Evaristo and McDonnell's calculation¹ implicitly assumes that Earth's entire landmass is forested, and leads to unrealistic results. For example, under Evaristo and McDonnell's median scenario¹, their table 1 implies that total post-clearing runoff in Asia would be 95% of total Asian precipitation⁸ (32,140 km³ yr⁻¹; Table 3), a runoff ratio that is rarely observed even in urban areas. For Australia and Oceania, the results in Evaristo and McDonnell's table 1 violate conservation of mass, with total post-clearing runoff (1,970 km³ yr⁻¹ + 5,412 km³ yr⁻¹ = 7,382 km³ yr⁻¹) exceeding total precipitation⁸ (6,405 km³ yr⁻¹).

Distributed over the roughly 40 million square kilometres of the Earth's surface that is actually forested⁷, Evaristo and McDonnell's

claimed global streamflow increase¹ of 34,098 km³ yr⁻¹ implies an average of 850 mm yr⁻¹ more streamflow from cleared forest lands. This value exceeds the streamflow increases that were measured in every one of the 95 paired watershed studies reviewed by Stednick⁹, and exceeds their average by a factor of five.

Back-of-the-envelope calculations suggest different conclusions. Globally, evapotranspiration from forests is roughly 250 mm yr⁻¹ greater than from croplands or grasslands¹⁰, and multiplying this difference by the 40 million square kilometres of global forests⁷ yields a rough estimate of 10,000 km³ yr⁻¹, less than one-third of Evaristo and McDonnell's¹ result. Even this may be an overestimate, because the lower evapotranspiration rates of grasslands partly reflect the fact that they often occur in drier climates; thus the difference between forest and grassland evapotranspiration may exaggerate the effects of converting forests to grasslands.

Concluding remarks

Evaristo and McDonnell are valued colleagues of ours, and we greatly appreciate their transparency in making their data and codes available, without which the issues described here would have been much harder to diagnose. We agree with them that streamflow response to forest management is an important issue that deserves a comprehensive analysis, including subsurface catchment characteristics as potential explanatory variables.

Readers should also keep in mind that this is not a purely academic exercise. How much, and under what conditions, forests should be cleared is an important policy question with wide-ranging consequences for economies, societies and ecosystems. In that regard, we are concerned that the conclusion that “forest removal can lead to increases in streamflow that are around 3.4 times greater than the mean annual

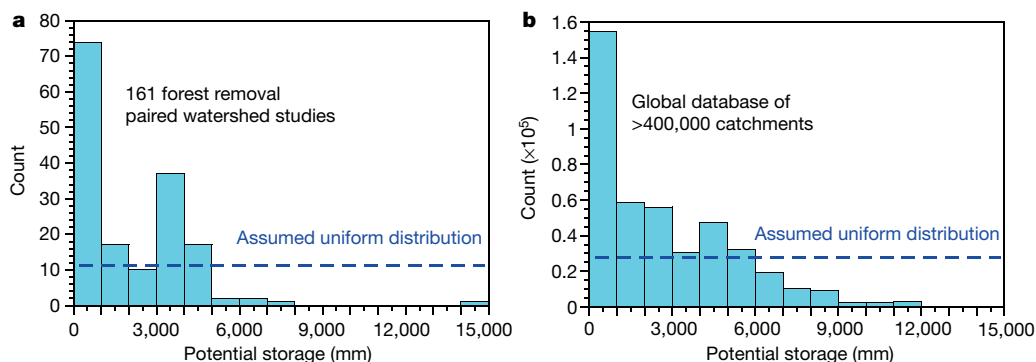


Fig. 2 | Distributions of potential storage, compared to the uniform distribution used to estimate its influence in Evaristo and McDonnell's analysis¹. **a**, Distribution of potential storage in Evaristo and McDonnell's

dataset of 161 paired watershed studies. **b**, Distribution of potential storage in Evaristo and McDonnell's database of over 400,000 catchments worldwide.

Table 3 | Modelled effects of forest cover change on continental runoff

Region	Total river runoff (km ³ yr ⁻¹) ^a	Change in runoff in response to forest-cover change ^a (km ³ yr ⁻¹)		Total river runoff after removal (km ³ yr ⁻¹) ^b	Total precipitation (km ³ yr ⁻¹) ^c	Change in runoff in response to forest-cover change (%) ^d		Median water yield in complete catchment dataset (%) ^e	
		Planting	Removal			Planting	Removal	Planting	Removal
Africa	4,320	-605(1,944)	8,986(5,616)	13,306	20,780	-14.0(45.0)	208.0(130.0)	-14(45)	208(130)
Asia	14,550	-1,979(5,835)	16,062(25,783)	30,612	32,140	-13.6(40.1)	110.4(177.2)	-14(40)	110(177)
Australia and Oceania	1,970	-412(725)	5,412(4,962)	7,382	6,405	-20.9(36.8)	274.7(251.9)	-21(36)	275(252)
Europe	3,240	-875(1,102)	813(1,426)	4,053	7,165	-27.0(34.0)	25.1(44.0)	-27(34)	25(44)
North and Central America	6,200	-806(2,034)	918(2,102)	7,118	13,910	-13.0(32.8)	14.8(33.9)	-13(33)	15(34)
South America	10,420	0(3,751)	1,908(17,559)	12,328	28,355	0.0(36.0)	18.3(168.5)	0(36)	18(168)
Totals	40,700	-4,676	34,098	74,799	109,755				

Values with parentheses are medians (and interquartile ranges).

^aFrom table 1 of ref. ¹.

^bSum of total river runoff and median change due to removal.

^cTotal precipitation from ref. ⁸, which is also the original source of the total river runoff values.

^dMedian and IQR of runoff changes, as percentage of total river runoff.

^eMedian and IQR of water yield predictions (each rounded to the nearest percentage point in the published database) for Evaristo and McDonnell's 442,319 'complete' catchments. These agree within roundoff error with the percentages calculated by dividing the change in runoff by the total runoff for each continent. This agreement demonstrates that the changes in runoff shown in table 1 of ref. ¹ were calculated by multiplying the median (and IQR) of the percentage water yield predictions by the total river runoff, rather than by the runoff from forested areas.

runoff of the Amazon River¹¹ is overstated and could be misinterpreted. The Amazon flows continuously, but the streamflow benefits of forest clearing are transient, typically lasting only a few years, or at most decades, after felling¹¹. One must also keep in mind that the water transpired by vegetation is an important source of precipitation farther downwind, estimated to account for roughly 40% of continental precipitation¹⁰. Thus, sustained large-scale clearing of forests would predictably lead to precipitation decreases and drying of continental interiors, although the precise magnitude of this effect remains difficult to constrain.

Data availability

All of the data analysed here are available as described in the data availability and code availability statements of ref. ¹, or from the cited references.

1. Evaristo, J. & McDonnell, J. J. Global analysis of streamflow response to forest management. *Nature* **570**, 455–461 (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019); Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019); Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).
2. Stoof, C. R. *et al.* Hydrological response of a small catchment burned by experimental fire. *Hydrol. Earth Syst. Sci.* **16**, 267–285 (2012).
3. Ruprecht, J. K. & Schofield, N. J. Effects of partial deforestation on hydrology and salinity in high salt storage landscapes. I. Extensive block clearing. *J. Hydrol.* **129**, 19–38 (1991).

4. Lawrence, P. A. & Thorburn, P. J. (eds) *Changes in Hydrology, Soil Fertility, and Productivity of Brigalow Catchments Following Clearing Project Report RQR89002* (Department of Primary Industries, Brisbane, 1989).
5. Thornton, C. M., Cowie, B. A., Freebairn, D. M. & Playford, C. L. The Brigalow catchment study. II. Clearing brigalow (*Acacia harpophylla*) for cropping or pasture increases runoff. *Aust. J. Soil Res.* **45**, 496–511 (2007).
6. SAS Institute Inc. *JMP 14 Profilers* (SAS Institute Inc., 2018).
7. *FAO Global Forest Resource Assessment 2015* (Food and Agriculture Organization of the United Nations, 2016).
8. L'vovich, M. I. & White, G. F. in *The Earth as Transformed by Human Action* (eds Turner, B. L. *et al.*) 235–252 (Cambridge Univ. Press, 1990).
9. Stednick, J. D. Monitoring the effects of timber harvest on annual water yield. *J. Hydrol.* **176**, 79–95 (1996).
10. Ellison, D., Futter, M. N. & Bishop, K. On the forest cover-water yield debate: from demand- to supply-side thinking. *Glob. Change Biol.* **18**, 806–820 (2012).
11. Andréassian, V. Waters and forests: from historical controversy to scientific debate. *J. Hydrol.* **291**, 1–27 (2004).

Author contributions All authors discussed the issues raised here, and contributed to the writing. J.W.K. analysed the data and led the writing effort.

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

Additional information

Correspondence and requests for materials should be addressed to J.W.K.

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.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Forest age and water yield

<https://doi.org/10.1038/s41586-020-1941-5>

Adriaan J. Teuling^{1*} & Anne J. Hoek van Dijke^{1,2,3}

Received: 28 June 2019

Accepted: 2 December 2019

Published online: 12 February 2020

Arising from: Evaristo, J. & McDonnell, J. J. *Nature* <https://doi.org/10.1038/s41586-019-1306-0> (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019);

Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019);

Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).

Planting and removal of forest affect average streamflow (also referred to as water yield), but there is ongoing debate as to what extent this long-term difference between precipitation and evapotranspiration is modulated by local conditions. A recent paper by Evaristo and McDonnell¹ introduces a conceptual vegetation-to-bedrock model to explain variability in reported streamflow responses to changes in forest cover based on an analysis of seven factors that describe climate, soil properties and catchment size. Their analysis excludes well known controls—such as the percentage of catchment area under change², forest type and time since afforestation—that we show here to be important. By excluding these primary controls, Evaristo and McDonnell risk attributing water yield response to co-varying secondary controls rather than to the underlying causes.

We illustrate the importance of the record length (or time since afforestation) using unique longterm measurements of water yield made under controlled conditions. At Castricum in The Netherlands, and St Arnold in Germany, two large lysimeters were planted with coniferous and deciduous trees in the 1940s and 1960s, respectively,

while reference conditions (bare soil and grassland, respectively) were maintained in an additional lysimeter. At both stations, strong, consistent and continuing declines in average water yield response were observed over averaging periods that ranged from several years up to the whole experiment duration (Fig. 1), coinciding with a steady increase in tree height and biomass^{3,4} and in spite of possible limitations in rooting depth. The declines follow an exponential decay (with a coefficient of determination of 0.91 or larger) with an e-folding time τ of 15 years and a stronger water yield response for coniferous forest than for deciduous forest. As a result, each individual lysimeter already covers a range in water yield response of 30% up to 70%, comparable to the total range reported by Evaristo and McDonnell across different watersheds¹. Similar response times were found for afforestation experiments in deciduous broadleaf forest in North Carolina in the USA⁵ and at the German lysimeter station of Britz-Eberswalde⁶, while analysis of longterm streamflow data in Sweden revealed similar strong effects of forest biomass and age⁷.

The record length of the studies used by Evaristo and McDonnell¹ varies considerably from 1 year to 75 years, but is mostly lower than the timescale of water yield response to forest growth of 15 years (Fig. 1). Therefore, it is likely that the values reported in studies with record lengths of up to once or even twice the e-folding time (15–30 years) are in fact highly sensitive to the length of their record. The mixing of data with variable record lengths could explain why Evaristo and McDonnell

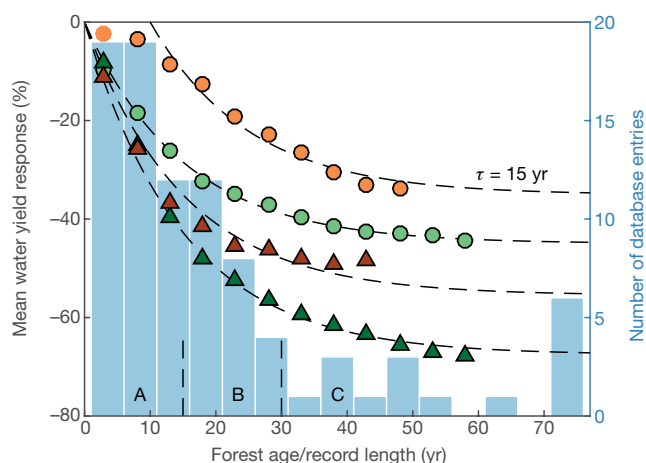


Fig. 1 | Impact of forest age on water yield response to forest planting. Data points are from coniferous (triangles) and deciduous (circles) lysimeters at Castricum (green) and St Arnold (red/orange). Dashed curves indicate exponential fits with a characteristic timescale τ of 15 years, with a 10-year shift assumed for the deciduous lysimeter in St Arnold. Letters A, B and C indicate record length (or forest age) domains used in Fig. 2. The background histogram shows the distribution of the record length of the forest planting studies used by Evaristo and McDonnell. Note that most studies (82%) have a record length of less than 30 years, and strong changes in water yield response are observed in this period. This figure and Fig. 2 were generated by Matlab 2015b (<http://nl.mathworks.com/products/matlab/>).

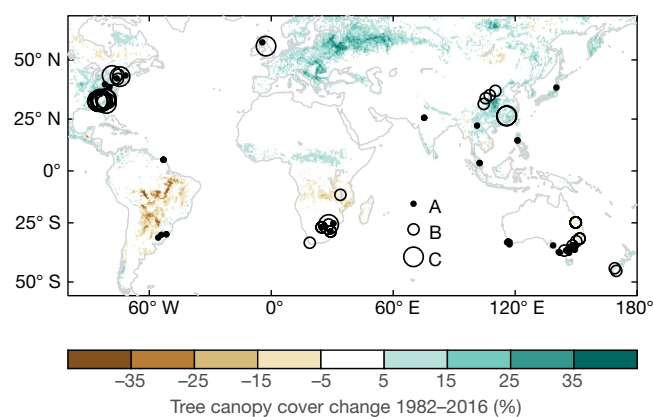


Fig. 2 | Global tree canopy cover change distribution and record length of water yield response to forest planting. Points/circles indicate locations of forest planting studies used by Evaristo and McDonnell¹, with the size reflecting the record length according to classes A, B and C as indicated in Fig. 1. The background map shows changes in tree canopy cover over the period 1982–2016 obtained from a recent analysis of satellite data⁸.

¹Hydrology and Quantitative Water Management Group, Wageningen University and Research, Wageningen, The Netherlands. ²Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research, Wageningen, The Netherlands. ³Environmental Sensing and Modelling, Environmental Research and Innovation Department, Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg. *e-mail: ryan.teuling@wur.nl

find actual evapotranspiration (AET) to be the factor explaining most of the magnitude, rather than timing, of water yield response to planting. When the location of stations with sufficient record length are added to a global map of changes in forest cover over the recent decades⁸, it becomes clear that accurate observations of long-term impacts of forest planting on water yield are concentrated in only a few regions. Strikingly, the forest cover change hotspots are observational blind spots for water cycle impacts. Given the potential of large-scale afforestation to offset carbon emissions⁹, a robust understanding of the hydrological impacts of current and future forest management is more important than ever.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Five-year-average water yield observations used in the analysis are provided in Extended Data Table 1.

1. Evaristo, J. & McDonnell, J. J. Global analysis of streamflow response to forest management. *Nature* **570**, 455–461 (2019); Addendum *Nature* <https://doi.org/10.1038/s41586-019-1586-4> (2019); Author Correction *Nature* <https://doi.org/10.1038/s41586-019-1588-2> (2019); Retraction *Nature* <https://doi.org/10.1038/s41586-020-1945-1> (2020).
2. Bosch, J. M. & Hewlett, J. D. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *J. Hydrol.* **55**, 3–23 (1982).
3. Harsch, N., Brandenburg, M. & Klemm, O. Large-scale lysimeter site St. Arnold, Germany: analysis of 40 years of precipitation, leachate and evapotranspiration. *Hydrol. Earth Syst. Sci.* **13**, 305–317 (2009).
4. Teuling, A. J. A forest evapotranspiration paradox investigated using lysimeter data. *Vadose Zone J.* **17**, 170031 (2018).
5. Swift, L. W. & Swank, W. T. Long term responses of streamflow following clearcutting and regrowth/Réactions à long terme du débit des cours d'eau après coupe et repeuplement. *Hydrol. Sci. Bull.* **26**, 245–256 (1981).
6. Müller, J. Forestry and water budget of the lowlands in northeast Germany—consequences for the choice of tree species and for forest management. *J. Water Land Dev.* **13a**, 133–148 (2009).
7. Jaramillo, F. et al. Dominant effect of increasing forest biomass on evapotranspiration: interpretations of movement in Budyko space. *Hydrol. Earth Syst. Sci.* **22**, 567–580 (2018).
8. Song, X.-P. et al. Global land change from 1982 to 2016. *Nature* **560**, 639–643 (2018).
9. Griscom, B. W. et al. Natural climate solutions. *Proc. Natl Acad. Sci. USA* **114**, 11645–11650 (2017).

Author contributions A.J.T. conceived the idea and carried out the analysis. A.J.H.v.d. contributed to interpretation of results and writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-1941-5>.

Correspondence and requests for materials should be addressed to A.J.T.

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.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Matters arising

Extended Data Table 1 | Observed water yield at long-term lysimeter stations

Site	Period	P	Reference	Broadleaf	Deciduous
St. Arnold	1966-1970	932.04	496.4	484.02	441.2
St. Arnold	1971-1975	677.28	357.24	340.44	191.8
St. Arnold	1976-1980	676.94	346.38	271.86	127.54
St. Arnold	1981-1985	773.34	439.62	334.94	198.8
St. Arnold	1986-1990	791.16	442.96	252.58	173.78
St. Arnold	1991-1995	872.9	530.08	328.16	276.72
St. Arnold	1996-2000	813.24	376.56	181.38	140.42
St. Arnold	2001-2005	835.98	391.74	153.74	171.48
St. Arnold	2006-2010	799.86	333.92	133.68	141.1
St. Arnold	2011-2013	703.43	253.57	130.6	NaN
Castricum	1941-1945	790.2	590.4	533.2	540.8
Castricum	1946-1950	791.4	596.4	433.4	351
Castricum	1951-1955	835.4	631.4	374	208.2
Castricum	1956-1960	857.6	664.4	339.4	190.6
Castricum	1961-1965	873.4	663.6	367.4	204
Castricum	1966-1970	910.8	700.2	366	175.6
Castricum	1971-1975	762.2	546	230.6	87.75
Castricum	1976-1980	783.6	597	270	122.8
Castricum	1981-1985	891.8	682	341.2	164.4
Castricum	1986-1990	848.8	657.2	361.2	118.4
Castricum	1991-1995	933.8	735.6	378.6	147.2
Castricum	1996-1997	744	550	145.5	2

Precipitation data are shown as reference. The reference lysimeter is grassland at St Arnold and bare soil at Castricum. Data after 2007 were not considered for the lysimeter with deciduous forest at St Arnold owing to storm damage caused by cyclone Kyrill. All units are millimetres per year.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- Data collection
- Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Comment on previous paper, with simple analysis of decades-old data that has been used in many previous studies
Research sample	Data comes from fixed lysimeters stations, one for each land cover type
Sampling strategy	No sampling involved
Data collection	AT the stations, observations have been done continuously for decades
Timing and spatial scale	Stations have operated for decades, size of each lysimeter approximately 400 m2
Data exclusions	Data after 2007 were not considered for the lysimeter with deciduous forest at St. Arnold due to storm damage caused by Cyclone Kyrill.
Reproducibility	n/a
Randomization	n/a
Blinding	n/a
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a
- Involvement in the study
- Antibodies
 - Eukaryotic cell lines
 - Palaeontology
 - Animals and other organisms
 - Human research participants
 - Clinical data

Methods

- n/a
- Involvement in the study
- ChIP-seq
 - Flow cytometry
 - MRI-based neuroimaging