

# SCIENTIFIC REPORTS



OPEN

## A deconvolutional Bayesian mixing model approach for river basin sediment source apportionment

William H. Blake<sup>1</sup>, Pascal Boeckx<sup>2</sup>, Brian C. Stock<sup>3</sup>, Hugh G. Smith<sup>4</sup>, Samuel Bodé<sup>2</sup>, Hari R. Upadhayay<sup>2,13</sup>, Leticia Gaspar<sup>5</sup>, Rupert Goddard<sup>1</sup>, Amy T. Lennard<sup>6</sup>, Ivan Lizaga<sup>5</sup>, David A. Lobb<sup>7</sup>, Philip N. Owens<sup>8</sup>, Ellen L. Petticrew<sup>8</sup>, Zou Zou A. Kuzyk<sup>9</sup>, Bayu D. Gari<sup>10</sup>, Linus Munishi<sup>11</sup>, Kelvin Mtei<sup>11</sup>, Amsalu Nebiyu<sup>10</sup>, Lionel Mabit<sup>12</sup>, Ana Navas<sup>5</sup> & Brice X. Semmens<sup>3</sup>

Increasing complexity in human-environment interactions at multiple watershed scales presents major challenges to sediment source apportionment data acquisition and analysis. Herein, we present a step-change in the application of Bayesian mixing models: Deconvolutional-MixSIAR (D-MIXSIAR) to underpin sustainable management of soil and sediment. This new mixing model approach allows users to directly account for the 'structural hierarchy' of a river basin in terms of sub-watershed distribution. It works by deconvoluting apportionment data derived for multiple nodes along the stream-river network where sources are stratified by sub-watershed. Source and mixture samples were collected from two watersheds that represented (i) a longitudinal mixed agricultural watershed in the south west of England which had a distinct upper and lower zone related to topography and (ii) a distributed mixed agricultural and forested watershed in the mid-hills of Nepal with two distinct sub-watersheds. In the former, geochemical fingerprints were based upon weathering profiles and anthropogenic soil amendments. In the latter compound-specific stable isotope markers based on soil vegetation cover were applied. Mixing model posterior distributions of proportional sediment source contributions differed when sources were pooled across the watersheds (pooled-MixSIAR) compared to those where source terms were stratified by sub-watershed and the outputs deconvoluted (D-MixSIAR). In the first example, the stratified source data and the deconvolutional approach provided greater distinction between pasture and cultivated topsoil source signatures resulting in a different posterior distribution to non-deconvolutional model (conventional approaches over-estimated the contribution of cultivated land to downstream sediment by 2 to 5 times). In the second example, the deconvolutional model elucidated a large input of sediment delivered from a small tributary resulting in differences in the reported contribution of a discrete mixed forest source. Overall D-MixSIAR model posterior distributions had lower (by ca 25–50%) uncertainty and quicker model run times. In both cases, the structured, deconvoluted output cohered more closely with field observations and local knowledge underpinning the need for closer attention to hierarchy in source and mixture terms in river basin source apportionment. Soil erosion and siltation challenge the energy-food-water-environment

<sup>1</sup>School of Geography, Earth and Environmental Sciences, University of Plymouth, Plymouth, UK. <sup>2</sup>Isotope Bioscience Laboratory – ISOFYS, Ghent University, Ghent, Belgium. <sup>3</sup>Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, USA. <sup>4</sup>Landcare Research, Palmerston North, New Zealand. <sup>5</sup>Soil and Water Department, Estación Experimental de Aula Dei (EEAD-CSIC), Zaragoza, Spain. <sup>6</sup>School of Environmental Sciences, University of Liverpool, Liverpool, UK. <sup>7</sup>Department of Soil Science, University of Manitoba, Winnipeg, Manitoba, Canada. <sup>8</sup>Quesnel River Research Centre, University of Northern British Columbia, Prince George, British Columbia, Canada. <sup>9</sup>Department of Geological Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. <sup>10</sup>College of Agriculture and Veterinary Medicine, Jimma University, Jimma, Ethiopia. <sup>11</sup>Nelson Mandela African Institute of Science and Technology, Arusha, Tanzania. <sup>12</sup>Soil and Water Management and Crop Nutrition Laboratory, Joint UN Food and Agricultural Organisation and International Atomic Energy Agency Division of Nuclear Techniques in Agriculture, Vienna, Austria. <sup>13</sup>Present address: Catchment Systems, Sustainable Agriculture Sciences, Rothamsted Research, North Wyke, Okehampton, UK. William H. Blake and Pascal Boeckx contributed equally. Correspondence and requests for materials should be addressed to W.H.B. (email: [william.blake@plymouth.ac.uk](mailto:william.blake@plymouth.ac.uk)) or P.B. (email: [pascal.boeckx@UGent.be](mailto:pascal.boeckx@UGent.be))

**nexus. This new tool for source apportionment offers wider application across complex environmental systems affected by natural and human-induced change and the lessons learned are relevant to source apportionment applications in other disciplines.**

**Context and aim.** Fingerprinting and (un)mixing concepts are used widely across environmental disciplines for forensic evaluation of pollutant sources<sup>1</sup>. In freshwater and marine systems, this includes tracking the source of organic and inorganic pollutants in water<sup>2–5</sup> and linking problematic sediment to soil erosion and specific land use sources<sup>6–10</sup>. It is, however, the complexity of ecological systems that has driven development of sophisticated Bayesian mixing models to appropriately represent inherent hierarchy and uncertainty in biogeochemical tracer data<sup>11</sup>.

In river basin sediment and contaminant mixing applications, the main parameters of interest are the proportions each ‘source’ contributes to a downstream ‘mixture’ within a river network wherein sources and mixtures are nested within the river basin and its sub-watershed structure. Knowledge of sediment source, transfer and residence time dynamics is critical to underpin sustainable land management for future food, water and energy security<sup>12</sup> particularly for vulnerable communities threatened by socio-economic impacts of soil erosion<sup>13,14</sup>. Soil erosion threatens food security<sup>15</sup> and associated siltation and pollution of river channels, lakes and reservoirs threatens river basin ecosystem service provision<sup>16,17</sup>, water security<sup>18</sup>, and hydro-electric power generation<sup>19</sup>. In response to challenges in tackling soil erosion and siltation problems worldwide, geochemical, radiochemical and isotopic fingerprinting techniques have developed considerably<sup>20</sup>. While these datasets capture real world sediment and pollutant source complexity in time and space, traditional statistical approaches used to select and treat these datasets can compromise source discrimination and apportionment. The power of geochemical, radiochemical and isotopic analytical techniques for sediment source and pollution apportionment may be more fully realized when these data are effectively coupled with new Bayesian modelling approaches.

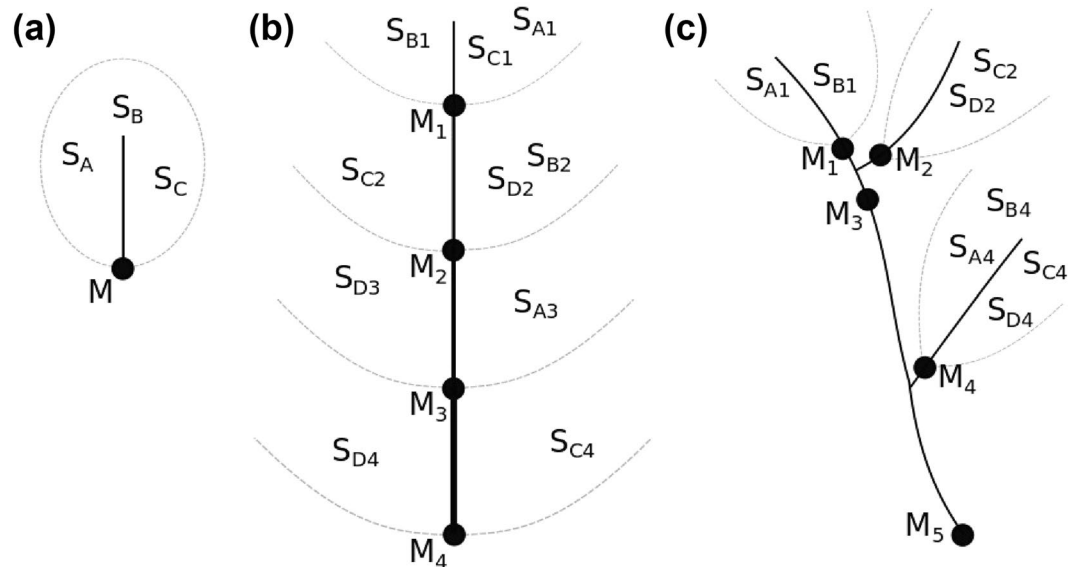
In environmental and ecological mixing problems, a key advantage of Bayesian over conventional linear mixing models is their flexible likelihood-based structure which permits better representation of inherent variability in source and mixture tracer data due to environmental processes<sup>21–23</sup>. Bayesian models also enable existing knowledge, in the form of ‘prior’ probability distributions, to be combined with new tracer data to obtain updated ‘posterior’ probability distributions for parameters of interest. To date these models have primarily been used in ecology to evaluate diet composition, population structure, and animal movement<sup>24–26</sup>. In this context, a new Bayesian mixing model framework MixSIAR<sup>11,27</sup>, was developed. MixSIAR is not a single model but a general framework that can create many different models based on available data types and elective parameterisation<sup>27</sup> with particular attention to the advantages of working with hierarchy in source and mixture data<sup>11,27</sup>.

While several river basin studies have successfully used versions of Bayesian mixing models to unmix sediment sources<sup>21,22,28–34</sup>, comparison of mixtures to sources is restricted to defined catchment units with potential loss of diagnostic detail with increasing scale. Here we present a step-change in application of Bayesian mixing models to river basin source apportionment problems with development of a hierarchical mixing model approach ‘*Deconvolutional MixSIAR*’ (D-MixSIAR). This extension of MixSIAR<sup>11</sup> allows users to directly account for the ‘structural hierarchy’ of a river basin or watershed. Accordingly, this contribution aims to demonstrate the application of the D-MixSIAR approach to both geochemical and compound specific stable isotope (CSSI) tracer data in lowland agricultural (UK) and upland forested (Nepal) watershed settings. Herein we aim to outline the advantages of a new approach to stratifying source apportionment data by watershed that offers unique insight into complex river basin process dynamics along the soil-sediment continuum.

**Source apportionment in river networks: a new deconvolution approach.** A key innovation of MixSIAR is the ability for the model to handle hierarchy in source and mixture data. In the river basin context, hierarchy is most obviously manifest in terms of river basin structure (e.g. nested or distributed watersheds from basin scale down in size to hillslope segments, Fig. 1). D-MixSIAR works by applying the MixSIAR mixing model sequentially to successive sediment mixture nodes (e.g. below a significant confluence) progressing downstream in a river network. Critically in D-MixSIAR, the source data for each node is stratified by sub-watershed. The full mathematical formulation details of MixSIAR are presented by Stock *et al.*<sup>11</sup>. Here we focus on the D-MixSIAR innovation and explain its application in the context of the two case studies. Full implementation code for D-MixSIAR is provided in Supplementary Information 1 (SI1).

A sediment mixture sampled from any node (with a target minimum three samples in total per node where feasible) in the river network is viewed as a mixture in relation to all upstream sources (Fig. 1a) and/or nodes (Fig. 1b and c), depending on location within the watershed structure. We propose two generic structural settings in which D-MixSIAR can be applied. Firstly, a longitudinal nested sub-watershed design (Fig. 1b) with, for example, four sources ( $S_A$ ,  $S_B$ ,  $S_C$ , and  $S_D$ ) sampled from hillslopes across the river basin wherein sampling is stratified across nested sub-watershed 1 to 4, and sediment mixtures (at nodes  $M_1$ ,  $M_2$ ,  $M_3$  and  $M_4$ ) are collected at the outlet of each nested sub-watershed. The mixtures can theoretically be unmixed against the sources pooled across the entire watershed or sources nested within a specific sub-watershed by node (i.e. source data structured by sub-watershed). Upstream mixtures can also be included as ‘sources’ to downstream mixtures (i.e.  $M_3$  unmixes via  $M_2$ ,  $S_{A3}$ ,  $S_{D3}$  in Fig. 1b). Alternatively, the hypothetical sources might be stratified across sub-watersheds which are distributed throughout the wider river basin (Fig. 1c) wherein all mixture nodes contribute to river basin outlet sediment.

There are hence two ways that sediment mixtures might be unmixed against source materials. First, if the sub-watershed structure is ignored as in previous work, sediment mixtures can be unmixed using a watershed-wide source signature from samples pooled across sub-watersheds (e.g. for source D in Fig. 1b,



**Figure 1.** Example experimental designs demonstrating how MixSIAR apportions sources in hierarchical river networks. In all designs, rivers flow downward, filled circles represent nodes at which sediment mixture (Mix = M) samples could be collected, and dashed grey lines delineate watersheds (S) denoted by subscript numbers, and subscript letters indicate unique sources. **(A)** Simple watershed with three sources,  $S_{A-C}$ , and one mixture location at the outflow, M. **(B)** Longitudinal system with four sources,  $S_{A-D}$ , and multiple mixture locations at the outlet of each nested subwatershed,  $M_{1-4}$ . **(C)** Distributed system with mixtures at the outflow of each of three subwatersheds,  $M_1$ ,  $M_2$ , and  $M_3$ , four sources ( $S_{A-D}$ ), as well as mixtures on the main channel:  $M_4$  and  $M_5$ . (note: not all sources are present in all subwatersheds).

combine  $S_{D2}$ ,  $S_{D3}$ , and  $S_{D4}$  into  $S_D$ , and so on for  $S_A$ ,  $S_B$  and  $S_C$ , and then unmix  $M_4$  against  $S_A$ ,  $S_B$ ,  $S_C$ ,  $S_D$ ). We refer to this as the “*pooled MixSIAR*” approach. Alternatively, the *D-MixSIAR* method can be applied to conduct a sequential analysis using stratified upstream mixtures as sources to downstream mixtures. The individual sub-watersheds are analysed with conventional MixSIAR and the results integrated to deconvolute the upstream mixtures at each level in terms of primary source contributions determined from the level above. For example, in Fig. 1b, this is manifest through (1) unmixing  $M_1$  against  $S_{A1}$ ,  $S_{B1}$ ,  $S_{C1}$ ; (2) unmixing  $M_2$  against  $M_1$ ,  $S_{B2}$ ,  $S_{C2}$ ,  $S_{D2}$ ; (3) unmixing  $M_3$  against  $M_2$ ,  $S_{A3}$ ,  $S_{D3}$ ; and finally (4)  $M_4$  against  $M_3$ ,  $S_{C4}$ ,  $S_{D4}$ . The *D-MixSIAR* routine (SI1) then recovers the overall contribution of the sources A-D to each of the mixtures by multiplying the estimated sub-watershed source proportions by the proportion each subwatershed contributes to the next mixture i.e. the proportion (p) of source B to mixture  $M_2$  is  $p_{B1} \cdot p_{M1} + p_{B2}$ . Importantly, *D-MixSIAR* propagates uncertainty in the proportion estimates, because it estimates full posterior distributions for each of the proportions. This principle can be extended to cover the nested scenario (Fig. 1c).

## Methods

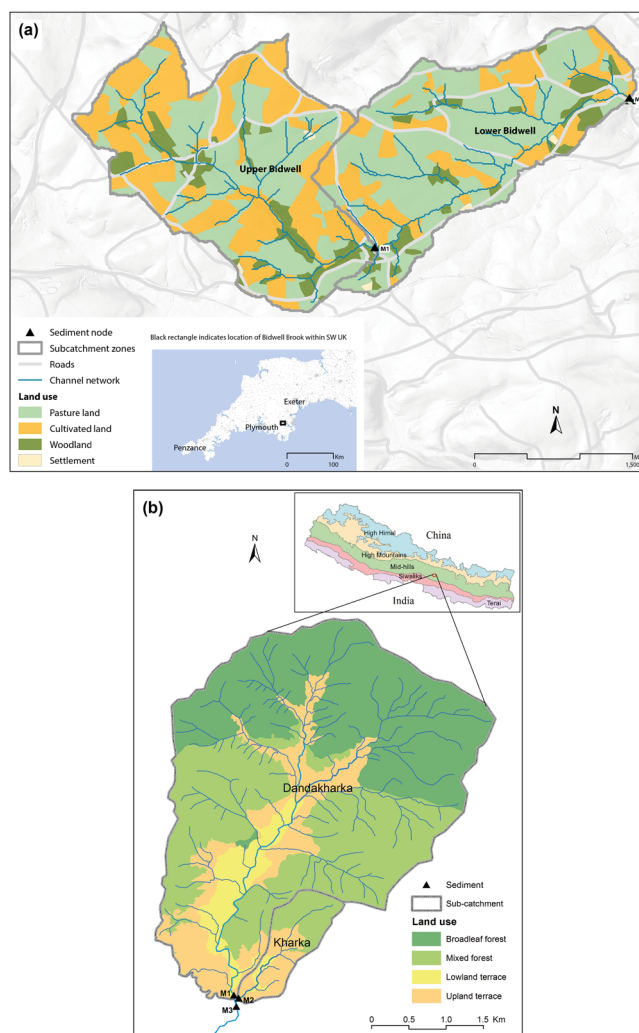
**Study watersheds.** To evaluate *D-MixSIAR* in the context of longitudinal (Fig. 1b) and distributed (Fig. 1c) watershed systems, source apportionment data were collected from (1) a lowland mixed agricultural catchment in southwest UK (Bidwell Brook) and (2) an upland mixed forest and agricultural watershed in Nepal (Upper Chitlang) (Fig. 2a and b). The former utilised major and minor element geochemistry, the latter used CSSIs<sup>35</sup>.

Bidwell Brook watershed (12 km<sup>2</sup>) has a maritime temperate climate receiving ca 1000 mm rainfall per annum with a notable wet autumn and winter period. Derived from a sedimentary substrate, the soil is a freely draining loam with moderate risk of damage by farm machinery (surface compaction and subsurface plough pan). Land use across the watershed is a mixture of pasture (P) and cultivated (CU) land (Table 1) where cultivated land is in rotation between i.e. maize, wheat, barley, root crops and ley grass. Upper and lower zones of the watershed have two notable differences in land management practice. Cultivation in the upper section of the study area is undertaken under certified organic practice i.e. no mineral fertiliser application and fields are generally buffered by grass strips. Pasture fields in the upper catchment are predominantly permanent (PP) on steeper slopes and used for grazing of dairy cattle with the remainder in rotation (RP). In the lower section of the watershed, cultivated land is farmed conventionally and the greater proportion of pasture fields is in rotation (RP).

The Upper Chitlang watershed receives ca 1600 mm rainfall per year 77% of which falls during a monsoon period from June to September. Soil is classified as a Cambisol that has developed over sedimentary alluvium substrate. In this study, two sub-watersheds that flow to a common tributary were selected, the Dandakharka (14 km<sup>2</sup>) and Kharka (1 km<sup>2</sup>) (Fig. 2). At the time of sampling, the Dandakharka had four dominant land uses (Table 1) i.e. broadleaf forest (BLF), mixed forest (MF), lowland agricultural terraces (LL) and upland agricultural terraces (UP). Rice is the dominant crop on lowland terraces during monsoon season with flood irrigation. Wheat and commercial vegetables are grown in the winter season. Upland terraces are rain fed with maize as dominant crop, intercropped with finger millet and vegetables. Upland terraces are more scattered and fragmented. Land

Watershed unit	Model Component	Source land cover (%)				Area (km <sup>2</sup> )
		P	CU	RM	CB	
Upper Bidwell	Mix 1	42	47	<1%	<1%	4.20
Lower Bidwell	Mix 2	51	38	<1%	<1%	7.90
		BLF	MF	LL	UP	
Dandakharka	Mix 1	41	38	6	16	14.4
Kharka	Mix 2	n/a	48	7	44	1.00
Upper Chitlang	Mix 3	38	36	6	19	15.4

**Table 1.** Sampled source distributions in (sub) watersheds of the Bidwell watershed (UK) and the Chitlang watershed (Nepal) where P = pasture, CU = cultivated land, RM = road-derived material and CB = channel bank, BLF = broad leaf forest, MF = mixed forest, LL = lowland terraces and UP = upland terraces. Mix 1–2 (Bidwell) and Mix 1–3 (Chitlang) refer to Fig. 1b and c respectively. Note the remainder of land cover in Bidwell Brook is largely stable woodland which was not included in the model.



**Figure 2.** Study watersheds. (a) Bidwell Brook, south west UK and (b) Upper Chitlang, Nepal where M1–M3 refers to the sediment mixture sampling nodes (see Fig. 1) and land use cover relates to identified sources.

use in the Kharka was similar but without BLF. Forest types were based on dominant vegetation but it should be noted that MF zones were formerly natural BLF so remnant and young broadleaf trees are also present. Leaf litter collection is common practice among farmers for livestock bedding material.

**Source and mixture sampling and analysis.** In both systems, sources were sampled in separate groups for each sub-watershed, i.e. sources were stratified. Depending on source extent, between 10 and 30 composite

samples (each a combination of ca. 15 random samples from the site) were collected from the upper 20 mm of soil at sites that represented the source class. For the Bidwell site, total source sample numbers were 32, 34 and 38 for CU, PP and RM respectively. For the Chitlang site, total source samples were 11, 13, 14 and 15 for BLF, LL, MF and UP respectively. The Bidwell study included samples of channel bank material scrapes ( $n = 32$ ) where exposed. Source materials were air dried ( $<45^{\circ}\text{C}$ ) and disaggregated. For CSSI analysis, samples were sieved to  $<2\text{ mm}$  as the signatures are independent of grain size<sup>35</sup>. For geochemical analysis, samples were sieved to  $<63\text{ }\mu\text{m}$  to minimise grain size effects on tracer signals<sup>36</sup>.

At each node (M1 and M2) in the longitudinal model of the Bidwell watershed, channel-bed deposited sediment<sup>37</sup> ( $n = 5$  and  $3$  for each node respectively) and time-integrated suspended sediment samples<sup>38</sup> ( $n = 1$  and  $3$ ) were collected through the autumn/winter period (access to suspended sediment samplers was restricted by high flow at node 1). Since we hoped to gain inference into potential contrasting source dynamics of these two sediment budget components, we parameterised sediment mixture ‘type’ (bed versus suspended) as a factor in MixSIAR<sup>11</sup> noting that while the time-integrated sampler numbers were limited, the nature of collection over a series of events underpinned representativeness. In the Chitlang distributed model, time-integrated suspended sediment samples<sup>38</sup> were collected at each of the three nodes ( $n = 9, 8$  and  $8$ ) where nodes M1 and M2 represented the sub-watersheds and node M3 represented the resulting mix in the higher order channel after the confluence. Time-integrated samples were collected across three periods in the wet season (i.e. early, mid and late), as we expected differences in sediment proportions due to seasonally-dependent land use practices. Since we hoped to gain inference on seasonal effects on the mixing process, we parameterised season as a factor. The “residual only” error formulation<sup>11,23</sup> was used due to limited mixture sample numbers at the nodes with application of factors, which is likely to be a common scenario in river basin applications. All sediment samples were dewatered through a combination of settling and centrifugation prior to freeze drying and disaggregation and sieving as above.

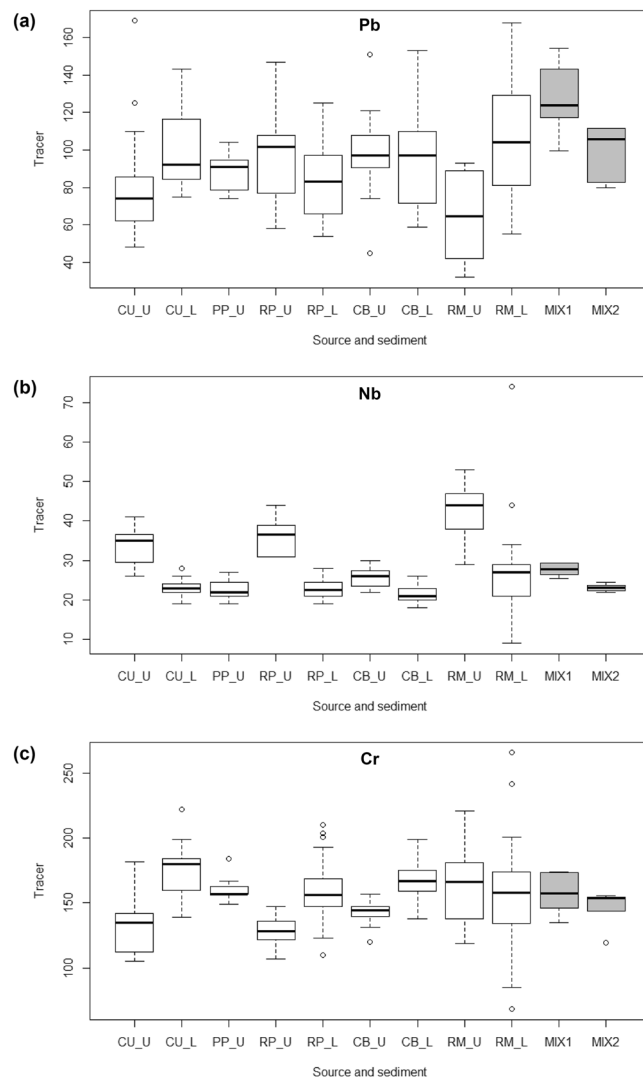
All samples from the Bidwell watershed were analysed for major and minor element geochemistry by Wave Length Dispersive X-Ray Fluorescence (WD-XRF; PANalytical Axios Max; OMNIAN application) as pressed pellets. All samples from the Chitlang watershed were prepared for CSSI analysis (Fatty Acids, FAs) by capillary gas chromatography-combustion-isotope ratio mass spectrometry (GC-C-IRMS; Trace GC Ultra interfaced via a GC/C III to DeltaPLUS XP, Thermo Scientific, Bremen, Germany) as described by Upadhayay *et al.*<sup>39</sup>.

**Tracer selection.** In light of Bayesian modelling advantages, we adopted a simplified tracer screening, i.e. range test, process. This step away from relying on ‘objective’ statistical techniques for selecting tracers was (1) based on recent proposals that a biogeochemical basis for selection is logically more appropriate<sup>40,41</sup> and (2) because the covariance structure of MixSIAR<sup>11</sup> handles redundancy so tracer selection by discriminant function analysis is not required, and might reduce discrimination or lead to erroneous outcomes. MixSIAR accounts for uncertainty in source and mixture data due to sampling or natural variability in the field. Two mechanisms accomplish this: fitting source tracer values within the model<sup>11,27</sup>, and specifying a distribution for the mixture data (i.e. an error structure<sup>23</sup>). In the Bayesian framework, striving to eliminate ‘redundant’ tracers is secondary to ensuring that tracers used are behaving independently and conservatively in the environment and indeed inclusion of even weak tracers can only improve model representation (cf.<sup>42</sup>). This differs from recent observations made using linear unmixing models<sup>43</sup> because of the way the model is formulated.

For all tracers, boxplots were produced for each set of sources and associated mixtures and mixture data assessed to see if they largely fell within or outside of the sources. Tracers that were clear poor performers, i.e. the mix values were largely outside the source range, were removed. Borderline tracers were retained based on the principle that Bayesian model convergence statistics<sup>25</sup> are the best assessment of fit. The approach implicitly assumed that sources and mixtures were representative and comparable and that correction factors for the tracers used were unnecessary<sup>36,44</sup>. MixSIAR assumes that mixture tracer data are normally distributed, which is appropriate because they are weighted combinations of the source means. Therefore, the central limit theorem applies, and a mixture of sources should be approximately normally distributed even if the sources are not<sup>11</sup>. Hence for the geochemical datasets, Exploratory Data Analysis (EDA) tools (histograms) were used instead of strict tests of normality. In addition, geochemical data that passed the range test were scrutinised for potential non-conservativeness in terms of fluvial sorting and biogeochemical process based on published data regarding environmental behaviour. For the CSSI datasets, basic descriptive statistics i.e. mean, and standard deviation (SD) were used to characterize the variation of  $\delta^{13}\text{C}$ -FA values within land uses and to describe data distributions. One-way analysis of variance (ANOVA) was used to assess the level of significance ( $p < 0.05$ ) of  $\delta^{13}\text{C}$ -FA ( $C_{22}$ - $C_{32}$ ) in the Chitlang watershed. Before ANOVA, all variables were checked for a normal distribution and homogeneity of variance. In case of significant ANOVA, means were compared by Tukey’s honestly significant differences (HSD) ( $p < 0.05$ ).

**Mixing model implementation.** MixSIAR is implemented as an open-source R package<sup>27</sup>. Full details of the mathematical formulation of MixSIAR are provided by Stock *et al.*<sup>11</sup>. Model performance has previously been validated<sup>23</sup> with simulation tests and 16 literature datasets, and the R package includes 11 examples with data that replicate published analyses<sup>27</sup>.

For the Bidwell study, MixSIAR was firstly formulated for each node separately with 15 tracers which passed the range test (Na, Mg, Al, P, S, Cl, Ca, Cr, Co, Cu, Ni, Ga, Rb, Nb, Ce) using a residual error term and sediment type as a factor and an uninformative prior (i.e. Dirichlet hyperparameters all set to 1 – see Stock *et al.*<sup>11</sup>). In a second model run we specified an informative prior wherein the Dirichlet hyperparameter for channel bank sources set at 0.01 because of potential overlap between channel bank and topsoil source signatures due to shared mixed weathering profiled origin<sup>44</sup> and limited field evidence of active bank retreat. The influence of a Dirichlet prior increases with fewer data points, greater source data variance, and poorer separation between source signatures<sup>11</sup>. For the Chitlang study, concentration-dependent<sup>45</sup> MixSIAR for each node was formulated with the



**Figure 3.** Example tracer distributions for (a) non-conservative (Pb), (b) weathering-controlled (Na) and (c) land management-amended (Cr) tracer properties in Bidwell Brook, in UK where Mix.1 and Mix.2 relate to sediment sampled at nodes M1 and M2 (See Fig. 1 and text for details). In box plots, median is shown by central line, interquartile range by box, range by whiskers with circles indicating outliers.

$\delta^{13}\text{C}$  values of six even FAs ( $\text{C}_{22}$ – $\text{C}_{32}$ ) for node M1 and three FAs ( $\text{C}_{24}$ ,  $\text{C}_{26}$  and  $\text{C}_{28}$ ) for node M2 using a residual error term, season as a factor and using an uninformative prior. The node posterior proportion contributions in both examples were then deconvoluted using the D-MixSIAR framework (section 2.1, SI1). For comparison to non-deconvolutional approaches, MixSIAR was subsequently formulated in both study watersheds using all source material signatures for each class pooled across both watershed units and all node mixtures from the systems compared to their respective pooled sources. Here the model was run with an uninformative prior (i.e. Dirichlet hyperparameters all set to 1).

In all MixSIAR model runs, the Markov Chain Monte Carlo (MCMC) parameters were generally set as follows: chain length = 1000000, burn = 700000, thin = 300, chains = 3. Convergence of all models was evaluated using the Gelman-Rubin diagnostic, rejecting model output if >5% of total variables was above 1.05, in which case chain length was increased.

## Results and Discussion

**Source tracer properties.** For the Bidwell case, of 26 major and minor geochemical elements measured in source and mixture samples of each node for D-MixSIAR, boxplot-based range testing (full data in Supplementary Information 1 (SI 1), examples in Fig. 3) led to elimination of seven elements (Pb, Ba, Br, Zn, Mn, Fe, Sr and Zr) based on lack of coherence between mix and source in either one or both nodes. Exclusion warrants some geochemical explanation<sup>46</sup>. Non conservative behaviour of Br can be explained by its propensity to form highly soluble salts<sup>47</sup>. Trace metals Pb and Zn concentrations were augmented in mixture sediment (Fig. 3a) suggesting an influence of sediment water interaction, noting high concentrations of these elements in road-related material) with a potential

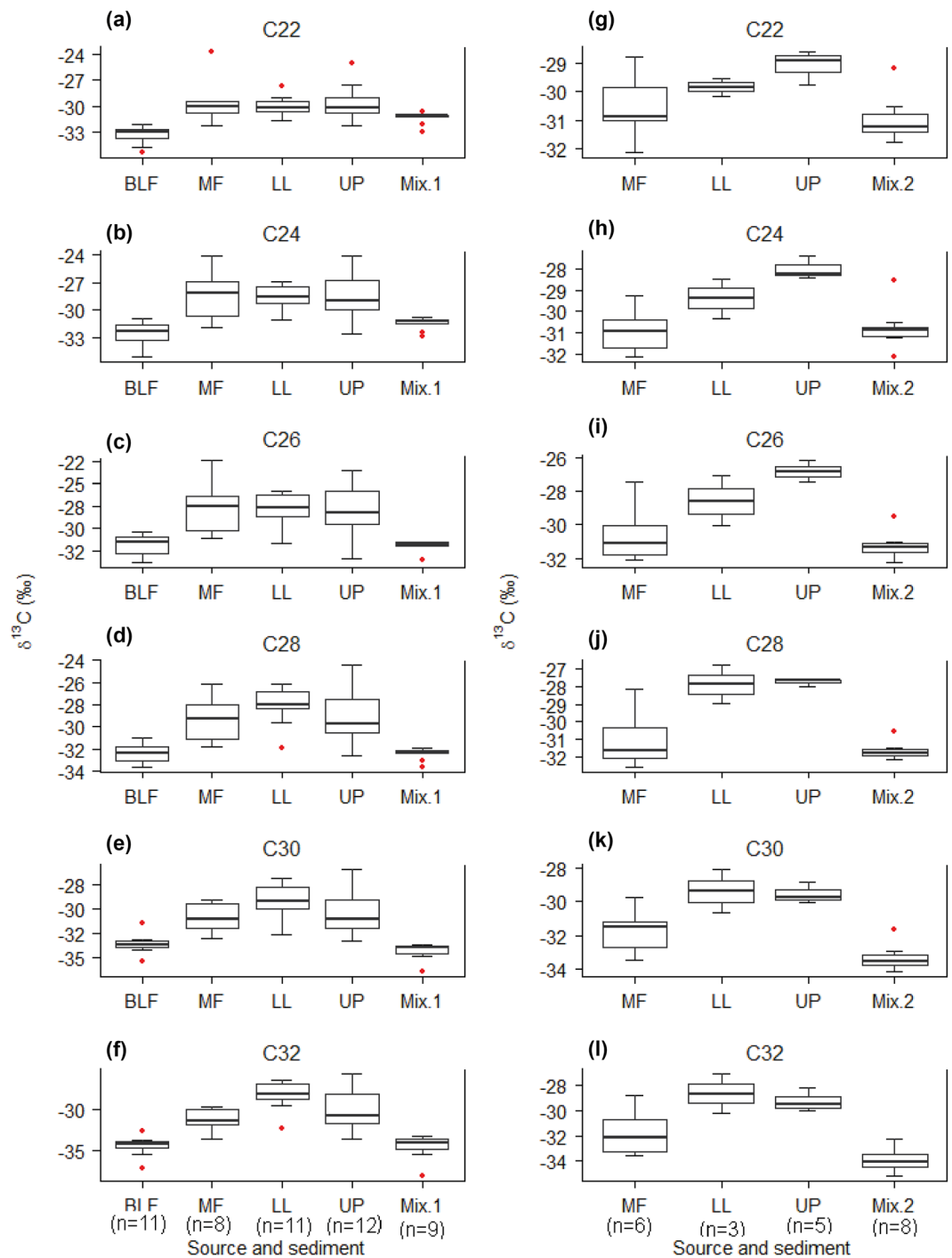
grain size control<sup>48</sup>. Barium, and Sr by association<sup>47</sup>, was also augmented in sediment either as a consequence of co-precipitation of weathering product with Fe and Mn<sup>47</sup> or mineralogical controls on composition due to preferential sorting. Iron and Mn also failed the range test implying a degree of environmental mobility and co-precipitation on fine particles<sup>48</sup>. Elimination of these range tested elements also raised concern regarding wider sorting, i.e. textural, controls on mineral composition i.e. changing proportions of silt *versus* clay minerals in mixtures which has been shown to exert a strong influence on sediment Fe, K, Si, Ti and Zr concentrations<sup>49</sup>. Given overlap with range test failures in Fe and Zr, these elements were also withdrawn on this basis. The viability of the remaining elements was underpinned by their known variability in the soil profiles due to weathering and mixing of soil horizons by cultivation (e.g. Na, Ca, Mg, Rb, Nb) (Fig. 3b) or association with anthropogenic amendments to soil<sup>50</sup> (e.g. Co, Cr, Cu) (Fig. 3c). For pooled MixSIAR, a wider range of mixture properties resulted in fewer exclusions: Ba, Br, Pb and Zn plus Fe, K, Ti and Si. Some elements retained might be considered less conservative in other systems (e.g. P, S and Cl). Here they were included due to clarity in discrimination between sources (SI 1) and limited scope for transformation at the scale of study<sup>51</sup>. Use of different elemental combinations had a notable outcome on posterior distributions and challenged convergence statistics, as discussed below.

Elimination of short-chain as well as unsaturated fatty acid (<20C atoms) tracers from the CSSI dataset was based on their biogeochemistry and behaviour in the soil and sediment environment<sup>39</sup>. These FAs are biosynthesised by both plants and microorganisms and are not stable in soil and sediment. In contrast, saturated long-chain FAs (>20C atoms) are exclusively biosynthesised by higher plants and are more stable than short-chained FAs<sup>52</sup>. Consequently,  $\delta^{13}\text{C}$  values of saturated long-chain FAs were hypothesised to characterise land uses defined by vegetation cover and it was assumed, based on the above, that they behaved conservatively in the watershed environment. Moreover, the range test was used to assess possible corruption in the isotopic value of FAs in the sediment (Fig. 4). In contrast to node M1, the  $\delta^{13}\text{C}$  values for node M2 mixtures fell outside the source range for chain length C<sub>22</sub>, C<sub>30</sub> and C<sub>32</sub>. This potentially implies poor characterisation of spatial variability of sources, especially in mixed forest due to inaccessibility or non-conservative behaviour. A variety of alternative factors are likely responsible for the variation of stable isotope signature of long-chain FAs including differences in the legacy of inputs from previous vegetation<sup>53</sup> and/or in current input from understory species<sup>54</sup>.

**Source apportionment: Deconvolutional MixSIAR versus Pooled MixSIAR.** *Longitudinal, lowland agricultural watershed.* D-MixSIAR and pooled MixSIAR posterior distributions, with the uninformative prior, showed differences in source apportionment for the Bidwell watershed (Table 2). In general, pooled MixSIAR apportionment data showed largely similar proportions across all of the identified factors i.e. node M1, node M2, bed sediment and suspended sediment with the exception of substantially lower channel bank contribution at node 1. Cultivated land was designated the dominant source in node 1 (51%) and node 2 (44%). In contrast, D-MixSIAR reported ca 10% cultivated soil input to node M1 and a greater proportion of pasture soil input (from both permanent and rotational pasture). Cultivated soil in the upper catchment is farmed under organic certification with extensive buffering that limits structural connectivity<sup>55</sup> to the stream. In contrast, steeper permanent pasture fields with notable poaching by livestock were observed to be strongly connected to the road network which drains directly into the brook at node M1. Additional structure in source data (i.e. separation of rotational and permanent pasture) led to tighter source terms and a different posterior outcome. Given the influence of farming practice on the cultivated soil signature in the upper watershed compared to the lower and the additional pasture categories when stratified by sub-watershed, we argue that D-MixSIAR gives a more credible posterior distribution in this case. Both pooled MixSIAR and D-MixSIAR posterior distributions imply a greater contribution of channel bank material to the bed sediment component in node M2 compared to suspended sediment which can be linked to localised channel bank slumping.

Previous studies have debated mixing model results regarding channel bank contributions<sup>56</sup> and the potential influence of overlapping source signatures<sup>44</sup>. The use of an informative prior based on observational evidence offers an additional analysis tool for users wherein the posterior output will represent an optimal compromise between the prior distribution and the data. When the arbitrary informative prior implying limited channel bank erosion was used, the proportion of channel bank material in model outputs (Table 3) was reduced by 10–20%. For node M1, bank contribution was reduced to zero in the D-MixSIAR posterior alongside a 10% increase in the topsoil sources. For node M2 channel bank contribution to suspended sediment was reduced to 4% in exchange for topsoil sources but for bed material the channel bank contribution was still notable at 18% (Table 3). Informative priors are more likely to be based on circumstantial evidence in river basin science as opposed to direct evidence as used in ecology (e.g. stomach contents informing animal diet analysis). In addition to observation-based inference, as in the above example, informative priors might readily be generated through (1) land use cover data overlain by erosion risk assessment<sup>57</sup>, (2) qualitative watershed walkover data<sup>58</sup>, (3) stream sediment load data, and (4) results from prior tracer studies<sup>59</sup>. Whatever users adopt to explore sensitivity of model output to prior information, it is critical that model outputs with uninformative and informative priors are compared and contrasted. In this case, while user inference was that channel bank erosion was minimal, model outputs demonstrated that this source was in fact an important contributor in the lowland reaches, especially to bed-stored material.

*Distributed upland forested watershed.* In the Nepalese forested catchment example, posterior distributions from MixSIAR run node-by-node (i.e. the raw ingredients of the D-MixSIAR prior to deconvolution) for node M1 (Table 3) implied that broadleaf forest contributed the majority (50–70%) of sediment to the Dandakharka stream network in all seasons, with a lesser but notable contribution from mixed forest (~20–35%) and the remainder from upland cultivation. In the Kharka subwatershed, posterior distributions implied that mixed forest dominated inputs to node M2 (Table 3). Raking leaf litter in the community managed forest is common practice<sup>60</sup> and is used as bedding material for livestock and we surmise that this disturbance of the forest floor litter and O-horizon leads to enhanced overland flow on the steep forest slopes. When node M3 was unmixed against nodes M1 and M2, the smaller system (Kharka) was shown to have contributed a disproportionate amount of the load



**Figure 4.** Distribution of  $\delta^{13}\text{C}$  (‰) values of FAs (C<sub>22</sub>–C<sub>32</sub>) in sources and sediments within Dandakharka (a–f) and Kharka (g–l) subcatchments (for M1 (Mix1) and M2 (Mix2) see detail in Fig. 2b). In box plots, median is shown by central line, interquartile range by box, range by whiskers with circles indicating outliers. Figures in the parenthesis in x-axis indicates the number of samples. Sources: broadleaf forest (BLF), mixed forest (MF), lowland terraces (LL) and upland terraces (UP).

(Table 3) for sub-watershed area i.e. ~25% of load from 6.5% of the total watershed. The Dandakharka watershed is less disturbed and larger than the Kharka.

Application of pooled MixSIAR to node M3 (Table 4) implied that broadleaf forest contributed the majority (~70–80%) of sediment to node M3 in all seasons, with minor contribution from mixed forest (~10–15%, Table 4). D-MixSIAR, however, based on deconvolution of the above individual model runs (Table 4) revealed clear differences in watershed-specific controls on sediment source dynamics because it took into account the contribution of each sub-watershed's stratified sources. Deconvolution of the posterior distributions gave substantially different results than the pooled MixSIAR approach. Broadleaf forest contributions to node M3 were



Model unit	Source	D-MixSIAR (factor = type)		Pooled-MixSIAR (factors = node and type)
		Susp. Sed	Bed sed.	
Upper Bidwell (Node M1)	CB	0.19 ± 0.13	0.14 ± 0.10	0.05 ± 0.06
	CU	0.11 ± 0.07	0.10 ± 0.07	0.51 ± 0.16
	PP	0.39 ± 0.13	0.47 ± 0.10	0.26 ± 0.15 (P)
	RP	0.15 ± 0.10	0.11 ± 0.07	
	RM	0.16 ± 0.05	0.17 ± 0.04	0.18 ± 0.11
Lower Bidwell (Node M2)	CB	0.22 ± 0.11	0.32 ± 0.09	0.36 ± 0.15
	CU	0.21 ± 0.10	0.19 ± 0.09	0.44 ± 0.15
	PP	0.27 ± 0.10	0.27 ± 0.07	0.09 ± 0.10 (P)
	RP	0.12 ± 0.04	0.13 ± 0.07	
	RM	0.18 ± 0.10	0.10 ± 0.03	0.11 ± 0.08
Bed sediment (both nodes M1 and M2)	CB	n/a	n/a	0.24 ± 0.17
	CU	n/a	n/a	0.39 ± 0.19
	P	n/a	n/a	0.16 ± 0.13
	RM	n/a	n/a	0.21 ± 0.16
Suspended sediment (both nodes M1 and M2)	CB	n/a	n/a	0.15 ± 0.13
	CU	n/a	n/a	0.40 ± 0.20
	P	n/a	n/a	0.25 ± 0.18
	RM	n/a	n/a	0.20 ± 0.15

**Table 2.** D-MixSIAR and pooled MixSIAR source apportionment data for the Bidwell watershed for (a) model runs with uninformative prior where CB is channel bank, CU is cultivated soil, PP is permanent pasture (node M1 only), RM is rotational pasture (i.e. cultivated at some point in the past), P is the former 2 combined (for pooled MixSIAR), and RM is material sampled from the roads. Note pooled MixSIAR delivers posterior distributions according to either node or type as factors.

Model unit	Source	D-MixSIAR (factor = type)		Pooled-MixSIAR (factors = node and type)
		Susp. Sed	Bed sed.	
Upper Bidwell (Node M1)	CB	0.00 ± 0.03	0.00 ± 0.02	0.02 ± 0.03
	CU	0.13 ± 0.08	0.13 ± 0.07	0.55 ± 0.19
	PP	0.51 ± 0.10	0.55 ± 0.08	0.24 ± 0.19 (P)
	RP	0.19 ± 0.11	0.15 ± 0.08	
	RM	0.16 ± 0.05	0.17 ± 0.04	0.19 ± 0.13
Lower Bidwell (Node M2)	CB	0.04 ± 0.06	0.18 ± 0.11	0.30 ± 0.15
	CU	0.24 ± 0.11	0.23 ± 0.14	0.51 ± 0.18
	PP	0.37 ± 0.13	0.34 ± 0.07	0.07 ± 0.12 (P)
	RP	0.21 ± 0.10	0.15 ± 0.08	
	RM	0.13 ± 0.04	0.11 ± 0.03	0.12 ± 0.09

**Table 3.** D-MixSIAR and pooled MixSIAR source apportionment data for the Bidwell watershed for model runs with an informative prior regarding limited channel bank input.

markedly reduced in favour of mixed forest contributions (Table 3). Despite both broadleaf and mixed forest areas being located on similar slope units, the mixed forest source was observed to have greater levels of disturbance due to proximity to settlements and its location in the drier, lower and more connected part of watershed (patches of forest fires were also observed in MF during field visit).

## Conclusion

In both the longitudinal example based on geochemical tracers and the distributed example based on compound specific stable isotope tracers, D-MixSIAR provided a different posterior distribution of sediment source contributions at the outlet of the defined study watersheds compared to conventional pooled MixSIAR. Field observations and local environmental knowledge suggest that the D-MixSIAR outputs are a more credible assessment of potential sediment sources in these systems supported by lower uncertainties and quicker model convergence.

In the longitudinal study, sediment yield from topsoil erosion was the dominant signal overall in both model outputs but it was only when specific topsoil sources were structured by watershed zone that more constrained source signatures emerged, i.e. organic versus conventional cultivation and permanent versus rotational pasture, permitting refined sub-watershed-specific results. In the distributed example, sediment yield from both forest types was important in pooled and D-MixSIAR and likely to be a function of surface erosion linked to disturbance of the soil during leaf litter collection by local communities and livestock trampling. Additionally,

Catchment	Source	Season		
		EW (April-June)	MW (July-Aug)	LW (Sep-Oct)
Dandakharka [Node M1]	BLF	0.70 ± 0.11	0.58 ± 0.12	0.50 ± 0.12
	MF	0.19 ± 0.1	0.28 ± 0.13	0.35 ± 0.14
	LL	0.03 ± 0.03	0.03 ± 0.04	0.03 ± 0.04
	UP	0.08 ± 0.08	0.11 ± 0.13	0.12 ± 0.14
Kharka [Node M2]	MF	0.75 ± 0.17	0.87 ± 0.14	0.76 ± 0.20
	LL	0.12 ± 0.12	0.06 ± 0.09	0.14 ± 0.17
	UP	0.13 ± 0.11	0.07 ± 0.10	0.09 ± 0.11
Confluence [Node M3]	Dandakharka	0.74 ± 0.17	0.83 ± 0.17	0.78 ± 0.2
	Kharka	0.26 ± 0.17	0.17 ± 0.17	0.22 ± 0.2
Model type	Sources	Season		
Deconvolutional MixSIAR [Node M3]	BLF	0.52 ± 0.14	0.41 ± 0.13	0.46 ± 0.15
	LL	0.05 ± 0.05	0.05 ± 0.06	0.04 ± 0.05
	MF	0.33 ± 0.13	0.42 ± 0.15	0.42 ± 0.17
	UP	0.09 ± 0.07	0.12 ± 0.13	0.09 ± 0.11
Pooled MixSIAR [Node M3]	BLF	0.79 ± 0.07	0.76 ± 0.08	0.81 ± 0.08
	LL	0.03 ± 0.03	0.03 ± 0.04	0.03 ± 0.03
	MF	0.12 ± 0.08	0.14 ± 0.10	0.11 ± 0.08
	UP	0.06 ± 0.05	0.07 ± 0.07	0.06 ± 0.06

**Table 4.** Temporal variability in relative contributions (mean ± SD) of sediment sources within individual sub-catchment and sub-catchments contribution to sediments downstream to confluence using MixSIAR node-by-node i.e. the raw ingredients of the D-MixSIAR prior to deconvolution. (Seasons: EW = early wet, LW = late wet and MW = mid-wet and sources: BLF = broad leaf forest, MF = mixed forest, LL = lowland, and UP = upland) (b) pooled MixSIAR versus D-MixSIAR relative contributions for node M3.

steep topography with dense tributaries and intersecting trackways means slopes are well connected to streams. D-MixSIAR, however, specifically distinguished broadleaf from mixed forests at the outlet because it was informed by the tributary proportions within the deconvolution process. Whereas pooled MixSIAR apparently overestimated broadleaf forest inputs, D-MixSIAR used stratified source samples to weight mixture proportions by sub-watershed; hence, downstream unmixing is constrained by results for upstream sub-watershed inputs, in the context of the key methodological assumptions articulated earlier.

In applications of D-MixSIAR, the source and mixture sampling strategy has to go hand-in-hand with the structural hierarchy of the drainage and sub-watershed network. A key assumption of the approach, in addition to the established need for source sample spatial representativeness, is that the sediment sampled at each node is spatially and temporally representative of the upstream contributing area for the time window of study i.e. the signature of sediment mixtures at a node may change with temporal dynamics of primary source contributions through a hydrological year. Apportionment to sub-watersheds is hence analogous to sediment load with an assumption of mass conservation in terms of source proportions to downstream mixtures. In this regard, user confidence in sample representativeness is paramount and sample numbers and/or mode of sample collection are critical considerations. The benefits of the stratified approach taken by D-MixSIAR in exploring complex systems introduces new sampling and analytical demands that users need to consider against specific research questions. As with other tracer studies, information of sediment load would be an advantage in interpreting the proportional data<sup>61,62</sup>.

We propose that the D-MixSIAR approach offers better system representation than current source apportionment approaches. The demonstration data presented, support distinct advantages of stratification of primary sources by sub-watershed within the D-MixSIAR model framework which (1) reduces the complexity of source groups contributing to high order stream mixture nodes by stratifying primary sources by sub-watershed, (2) decreases variability within sources by removing the duplication effect of tracer values and potential overlapping of source signatures and (3) appropriately weights sub-watershed specific primary source contributions by systematically accounting for sub-watershed contributions. In conventional mixing models at the larger watershed scale (analogous to pooled MixSIAR), having multiple sources with overlapping source tracer composition significantly increases model output uncertainty. The step-change proposed here in stratifying data according to watershed hierarchy, followed by deconvolution at the next level, clarifies this and leads to a model structure that represents better the soil-sediment continuum.

The approach further highlights the need and advantages of proper attention to data hierarchy in river basin systems. This new tool for source apportionment offers wider application across complex environmental systems affected by natural and human-induced change and the lessons learned are relevant to source apportionment applications in other disciplines. The new data handling routines provide evidence to support management of complex human-environment interactions and help tackle the global challenge of improved food, water and energy security.

### Data Availability

All data generated or analysed during this study are included in this published article and its Supplementary Information files: Datafile text and model code for demonstration examples.

## References

- Owens, P. N. *et al.* Fingerprinting and tracing the sources of soils and sediments: Earth and ocean science, geoarchaeological, forensic, and human health applications. *Earth-Science Rev.* **162**, 1–23 (2016).
- Baker, A., Inverarity, R., Charlton, M. & Richmond, S. Detecting river pollution using fluorescence spectrophotometry: case studies from the Ouseburn, NE England. *Environ. Pollut.* **124**, 57–70 (2003).
- Frank, R. A. *et al.* Profiling Oil Sands Mixtures from Industrial Developments and Natural Groundwaters for Source Identification. *Environ. Sci. Technol.* **48**, 2660–2670 (2014).
- Warner, N. R. *et al.* New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations. *Environ. Sci. Technol.* **48**, 12552–12560 (2014).
- Goody, D. C. *et al.* Isotopic Fingerprint for Phosphorus in Drinking Water Supplies. *Environ. Sci. Technol.* **49**, 9020–9028 (2015).
- Collins, A. L. *et al.* Sediment source fingerprinting as an aid to catchment management: A review of the current state of knowledge and a methodological decision-tree for end-users. *J. Environ. Manage.* **194**, 86–108 (2017).
- Walling, D. E. The evolution of sediment source fingerprinting investigations in fluvial systems. *J. Soils Sediments* **13**, 1658–1675 (2013).
- Haddadchi, A., Ryder, D. S., Evrard, O. & Olley, J. Sediment fingerprinting in fluvial systems: Review of tracers, sediment sources and mixing models. *Int. J. Sediment Res.* **28**, 560–578 (2013).
- Mukundan, R., Walling, D. E., Gellis, A. C., Slattery, M. C. & Radcliffe, D. E. Sediment Source Fingerprinting: Transforming From a Research Tool To a Management Tool 1. **48** (2012).
- Davis, C. M. & Fox, J. F. Sediment Fingerprinting: Review of the Method and Future Improvements for Allocating Nonpoint Source Pollution. *J. Environ. Eng.* **135**, 490–504 (2009).
- Stock, B. C. *et al.* Analyzing mixing systems using a new generation of Bayesian tracer mixing models. *PeerJ* **6**, e5096, <https://doi.org/10.7717/peerj.5096> (2018).
- Liang, L. K., Heng, K., Sakadevan, K., Dercon, G. & Nguyen, M. L. *Managing soils for food security and climate change adaptation and mitigation.* (Food and Agriculture Organisation of the United Nations, 2014).
- Wilson, G., Quaranta, G., Kelly, C. & Salvia, R. Community resilience, land degradation and endogenous lock-in effects: evidence from the Alento region, Campania, Italy. *J. Environ. Plan. Manag.* **59**, 518–537 (2016).
- Kelly, C. *et al.* Community resilience and land degradation in forest and shrubland socio-ecological systems: Evidence from Gorgoglione, Basilicata, Italy. *Land use policy* **46**, 11–20 (2015).
- Pimentel, D. Soil Erosion: A Food and Environmental Threat. *Environ. Dev. Sustain.* **8**, 119–137 (2006).
- Gilvear, D. J., Spray, C. J. & Casas-Mulet, R. River rehabilitation for the delivery of multiple ecosystem services at the river network scale. *J. Environ. Manage.* **126**, 30–43 (2013).
- Adel, M. M. Downstream ecocide from upstream water piracy. *Am. J. Environ. Sci.* **8**, 528–548 (2012).
- Ryken, N. *et al.* Impact of papyrus wetland encroachment on spatial and temporal variabilities of stream flow and sediment export from wet tropical catchments. *Sci. Total Environ.* **511**, 756–766 (2015).
- Syvitski, J. P. M. Supply and flux of sediment along hydrological pathways: research for the 21st century. *Glob. Planet. Change* **39**, 1–11 (2003).
- Dercon, G. *et al.* Fallout radionuclide-based techniques for assessing the impact of soil conservation measures on erosion control and soil quality: an overview of the main lessons learnt under an FAO/IAEA Coordinated Research Project. *J. Environ. Radioact.* **107**, 78–85 (2012).
- Cooper, R. J. *et al.* Apportioning sources of organic matter in streambed sediments: An integrated molecular and compound-specific stable isotope approach. *Sci. Total Environ.* **520**, 187–197 (2015).
- Cooper, R. J. & Krueger, T. An extended Bayesian sediment fingerprinting mixing model for the full Bayes treatment of geochemical uncertainties. *Hydrol. Process.* **31**, 1900–1912 (2017).
- Stock, B. C. & Semmens, B. X. Unifying error structures in commonly used biotracer mixing models. *Ecology* **97**, 2562–2569 (2016).
- Moore, J. W. & Semmens, B. X. Incorporating uncertainty and prior information into stable isotope mixing models. *Ecol. Lett.* **11**, 470–480 (2008).
- Semmens, B. X., Ward, E. J., Moore, J. W. & Darimont, C. T. Quantifying Inter- and Intra-Population Niche Variability Using Hierarchical Bayesian Stable Isotope Mixing Models. *Plos One* **4**, e6187 (2009).
- Parnell, A. C. *et al.* Bayesian stable isotope mixing models. *Environmetrics* **24**, 387–399 (2013).
- Stock, B. C. & Semmens, B. X. *MixSTAR GUI User Manual v3.1*, <https://doi.org/10.5281/zenodo.47719> (2017).
- Abban, B. *et al.* An enhanced Bayesian fingerprinting framework for studying sediment source dynamics in intensively managed landscapes. *Water Resour. Res.* **52**, 4646–4673 (2016).
- Massoudieh, A., Gellis, A., Banks, W. S. & Wiczorek, M. E. Suspended sediment source apportionment in Chesapeake Bay watershed using Bayesian chemical mass balance receptor modeling. *Hydrol. Process.* **27**, 3363–3374 (2013).
- D’Haen, K. *et al.* Unravelling changing sediment sources in a Mediterranean mountain catchment: A Bayesian fingerprinting approach. *Hydrol. Process.* **27**, 896–910 (2013).
- Dutton, C., Anisfeld, S. C. & Ernstberger, H. A novel sediment fingerprinting method using filtration: Application to the Mara River, East Africa. *J. Soils Sediments* **13**, 1708–1723 (2013).
- Koiter, A. J. *et al.* Investigating the role of connectivity and scale in assessing the sources of sediment in an agricultural watershed in the Canadian prairies using sediment source fingerprinting. *J. Soils Sediments* **13**, 1676–1691 (2013).
- Nosrati, K., Govers, G., Semmens, B. X. & Ward, E. J. A mixing model to incorporate uncertainty in sediment fingerprinting. *Geoderma* **217–218**, 173–180 (2014).
- Stewart, H. A., Massoudieh, A. & Gellis, A. Sediment source apportionment in Laurel Hill Creek, PA, using Bayesian chemical mass balance and isotope fingerprinting. *Hydrol. Process.* **29**, 2545–2560 (2015).
- Gibbs, M. M. Identifying Source Soils in Contemporary Estuarine Sediments: A New Compound-Specific Isotope Method. *Estuaries and Coasts* **31**, 344–359 (2008).
- Lacey, J. P. *et al.* The challenges and opportunities of addressing particle size effects in sediment source fingerprinting: A review. *Earth-Science Rev.* **169**, 85–103 (2017).
- Lambert, C. P. & Walling, D. E. Measurement of channel storage of suspended sediment in a gravel-bed river. *CATENA* **15**, 65–80 (1988).
- Phillips, J. M., Russell, M. A. & Walling, D. E. Time-integrated sampling of fluvial suspended sediment: a simple methodology for small catchments. *Hydrol. Process.* **14**, 2589–2602 (2000).
- Upadhyay, H. R. *et al.* Methodological perspectives on the application of compound-specific stable isotope fingerprinting for sediment source apportionment. *J. Soils Sediments* **17**, 1537–1553 (2017).
- Koiter, A. J., Owens, P. N., Petticrew, E. L. & Lobb, D. A. The behavioural characteristics of sediment properties and their implications for sediment fingerprinting as an approach for identifying sediment sources in river basins. *Earth-Science Rev.* **125**, 24–42 (2013).
- Lacey, J. P. & Olley, J. An examination of geochemical modelling approaches to tracing sediment sources incorporating distribution mixing and elemental correlations. *Hydrol. Process.* **29**, 1669–1685 (2015).
- Sherriff, S. C., Franks, S. W., Rowan, J. S., Fenton, O. & Ó’hUallacháin, D. Uncertainty-based assessment of tracer selection, tracer non-conservativeness and multiple solutions in sediment fingerprinting using synthetic and field data. *J. Soils Sediments* **15**, 2101–2116 (2015).

43. Pulley, S., Foster, I. & Antunes, P. The uncertainties associated with sediment fingerprinting suspended and recently deposited fluvial sediment in the Nene river basin. *Geomorphology* **228**, 303–319 (2015).
44. Smith, H. G. & Blake, W. H. Sediment fingerprinting in agricultural catchments: A critical re-examination of source discrimination and data corrections. *Geomorphology* **204**, 177–191 (2014).
45. Upadhayay, H. R. *et al.* Isotope mixing models require individual isotopic tracer content for correct quantification of sediment source contributions. *Hydrol. Process.* **32**, 981–989 (2018).
46. Smith, H. G., Karam, D. S. & Lennard, A. T. Evaluating tracer selection for catchment sediment fingerprinting. *J. Soils Sediments* **18**, 3005–3019 (2018).
47. Salminen, R., Tarvainen, T. & De Vos, W. Forum of the European Geological Surveys Directors & Geologian tutkimuskeskus (Finland). *Geochemical atlas of Europe*. (Geological Survey of Finland, 2005).
48. Horowitz, A. J. & Horowitz, A. J. *A primer on sediment-trace element chemistry*. (Lewis Publishers, 1991).
49. Cuvén, S., Francus, P. & Lamoureux, S. F. Estimation of grain size variability with micro X-ray fluorescence in laminated lacustrine sediments, Cape Bounty, Canadian High Arctic. *J. Paleolimnol.* **44**, 803–817 (2010).
50. Nziguheba, G. & Smolders, E. Inputs of trace elements in agricultural soils via phosphate fertilizers in European countries. *Sci. Total Environ.* **390**, 53–57 (2008).
51. Blake, W. H., Walsh, R. P. D., Sayer, A. M. & Bidin, K. Quantifying fine-sediment sources in primary and selectively logged rainforest catchments using geochemical tracers. *Interact. Between Sediments Water* **1**, 251–259 (2006).
52. Van Der Voort, T. S. *et al.* Diverse Soil Carbon Dynamics Expressed at the Molecular Level. *Geophysical Research Letters* **44**, 11,840–11,850 (2017).
53. Mabit, L. *et al.* Novel application of Compound Specific Stable Isotope (CSSI) techniques to investigate on-site sediment origins across arable fields. *Geoderma* **316**, 19–26 (2018).
54. Schäfer, I. K. *et al.* Leaf waxes in litter and topsoils along a European transect. *SOIL* **2**, 551–564 (2016).
55. Bracken, L. J. *et al.* Concepts of hydrological connectivity: Research approaches, pathways and future agendas. *Earth-Science Rev.* **119**, 17–34 (2013).
56. Collins, A. L., Walling, D. E., Webb, L. & King, P. Apportioning catchment scale sediment sources using a modified composite fingerprinting technique incorporating property weightings and prior information. *Geoderma* **155**, 249–261 (2010).
57. Wynants, M., Solomon, H., Ndaikidemi, P. & Blake, W. H. Pinpointing areas of increased soil erosion risk following land cover change in the Lake Manyara catchment, Tanzania. *Int. J. Appl. Earth Obs. Geoinf.* **71**, 1–8 (2018).
58. Thompson, J., Cassidy, R., Doody, D. G. & Flynn, R. Predicting critical source areas of sediment in headwater catchments. *Agric. Ecosyst. Environ.* **179**, 41–52 (2013).
59. Barthod, L. R. M. *et al.* Selecting Color-based Tracers and Classifying Sediment Sources in the Assessment of Sediment Dynamics Using Sediment Source Fingerprinting. *J. Environ. Qual.* **44**, 1605 (2015).
60. Baral, S. K. & Katzensteiner, K. Impact of biomass extraction on soil properties and foliar nitrogen content in a community forest and a semi-protected natural forest in the central mid-hills of Nepal. *Trop. Ecol.* **56**, 323–333 (2015).
61. Smith, H. G., Sheridan, G. J., Lane, P. N. J., Noske, P. J. & Hejnis, H. Changes to sediment sources following wildfire in a forested upland catchment, southeastern Australia. *Hydrol. Process.* **25**, 2878–2889 (2011).
62. Blake, W. H., Ficken, K. J., Taylor, P., Russell, M. A. & Walling, D. E. Tracing crop-specific sediment sources in agricultural catchments. *Geomorphology* **139**, 322–329 (2012).

## Acknowledgements

Research was supported by a European Commission Horizon 2020 programme under the Marie Skłodowska-Curie Research and Innovation Staff Exchange programme: IMIXSED project: *Integrating isotopic techniques with Bayesian modelling for improved assessment and management of global sedimentation problems* (Project ID 644320). The work represents a contribution to the IAEA Coordinated Research Project (CRP) D15017: *Nuclear Techniques for a Better Understanding of the Impact of Climate Change on Soil Erosion in Upland Agro-ecosystems*. The development of case study data was underpinned by the valued contributions of Prof. G.E. Millward, Dr Alex Taylor and Mr Maarten Wynants in the University of Plymouth Consolidated Radioisotope Facility. Shaun Lewin and Tim Absalom kindly assisted with production of Figures.

## Author Contributions

W.B., P.B., H.S. and A.N. developed the core concept of the article from the applied river basin perspective in collaboration with BSe and BSt who led the model development components. S.B. and H.U. and L.M., K.M. and B.G. led the demonstration dataset acquisition and analysis. L.G., R.G., A.L., I.L., D.L., P.O., E.P., Z.K., B.G., L.M., K.M., A.N. were active participants, with the senior authors, in workshops contributing to method development. L.M. provided steer on application needs and challenges.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-30905-9>.

**Competing Interests:** The authors declare no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018