Geoderma 268 (2016) 108–118

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Using multiple composite fingerprints to quantify fine sediment source contributions: A new direction

X.C. (John) Zhang [a,*], B.L. Liu [b,1]

[a] USDA-ARS Grazinglands Research Lab., El Reno, OK, USA
[b] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, Gansu, China

ARTICLE INFO

Article history:
Received 5 October 2015
Received in revised form 9 January 2016
Accepted 20 January 2016
Available online 5 February 2016

Keywords:
Sediment source fingerprinting
Tracer selection
Composite fingerprints
Mixing model
Model optimization

ABSTRACT

Sediment source fingerprinting provides an essential means for estimating sediment source contributions, which are needed not only for soil conservation planning but also for erosion model evaluation/refinement. A single optimum composite fingerprint has been widely used in the literature to estimate sediment provenance. The objectives of this work are to (1) verify whether an optimum composite fingerprint exists, (2) present a new direction of using multiple composite fingerprints to improve the accuracy and reliability of source contribution estimation, and (3) evaluate the optimization model formulation and the validity of the tracer discriminatory weighting. This study shows that tracer selection greatly impacts the estimated source contributions. The optimum composite fingerprint may not exist, or at least cannot be identified simply based on ability of the tracer to discriminate sources because of the lack of correlation between ability of the tracer to discriminate and its rigor in estimating source contributions. The weak link is likely caused by (1) tracer conflicts, (2) differential tracer measurement errors, and (3) varying degree of the conservativeness of each tracer or lack of it. To overcome this shortcoming, a new approach of using multiple composite fingerprints was proposed. Contrary to the popular tracer reduction strategies, this new approach uses a maximum number of composite fingerprints, which contain non-contradictory tracers in each composite, to maximize the use of all tracer information. The new approach assumes that source proportions averaged over multiple composite fingerprints are more likely to be closer to the population means than any estimate using a single fingerprint alone. Such a 'mean of the means' approach has been shown to not only improve the accuracy but also reduce the uncertainty of the proportion estimates. The model of the absolute relative difference performed slightly better than the squared relative difference model in estimating source contributions, suggesting the former be preferred. The results also indicated that the tracer discriminatory weighting should be excluded as it tends to bias contribution estimates.

1. Introduction

Quantitative information of sediment provenance is necessary to ensure that the best erosion control measures are placed in areas that maximize sediment reduction. For soil erosion control, the most effective measures should be placed where the most severe erosion occurs in a watershed. Apart from the conservation planning, knowledge of sediment sources (e.g., gully versus overland erosion) is needed for better calibrating and validating process-based hydrological and erosion models at a watershed scale such as Water Erosion Prediction Project (WEPP) model (Flanagan and Nearing, 1995), as many physically based models simulate each erosion process or erosion type separately. However, measured data on sediment contributions from different sediment source types or erosion processes are extremely limited, simply because many sediment source types or sediment generated from different erosion processes cannot be readily or directly measured in a watershed setting. Generally there are two approaches in measuring sediment provenance. One is to use erosion pins or unit source plots or small watersheds to measure sediment provenance from a particular source type such as gully or overland erosion source. This approach is often restricted to small areas so that a particular sediment source can be isolated and directly measured. Another approach is to estimate sediment source contributions using various sediment source fingerprints. The fingerprinting approach is applicable to any spatial scale, ranging from a plot to a large river basin (e.g., Zhang et al., 2003; Walling et al., 1993; Collins et al., 1997; Koiter et al., 2013; Laceby and Olley, 2014).

Different fingerprint properties have been used to discriminate among potential sediment sources, including physical, geochemical, isotopical, and biological properties. A single property such as ^{137}Cs has been successfully used to distinguish between two sediment

Abbreviations: DFA, discriminant function analysis; GA, genetic algorithm; GOF, goodness of fit; MAF, mean absolute fit; MC, Monte Carlo; PCA, principal component analysis; RD, relative difference.

* Corresponding author at: 7207 W. Cheyenne St., El Reno, OK 73036, USA.
E-mail address: john.zhang@ars.usda.gov (X.C.(J.) Zhang).
[1] 320 Donggangxi Road, Lanzhou 730000, Gansu, China.

http://dx.doi.org/10.1016/j.geoderma.2016.01.031
0016-7061/Published by Elsevier B.V.

sources (Olley et al., 1993; Wallbrink et al., 1996; Nagle and Ritchie, 1999, 2004; Nagle et al., 2007; Ritchie et al., 2009) or potentially three sources (Mukundan et al., 2010). However, as the number of sources increases, a single property becomes inadequate to discriminate a wide range of potential sources, and thus a composite fingerprint of multiple properties has been used (e.g., Walling et al., 1993; Walling and Woodward, 1995; Collins et al., 1997, 1998, 2010; Collins and Walling, 2002; Walling, 2005; Mukundan et al., 2010; Wilkinson et al., 2013; Koiter et al., 2013; Laceby and Olley, 2014). Tracer selection is crucial in sediment source fingerprinting, because different tracers may predict very different proportional contributions (Sherriff et al., 2015; Laceby et al., 2015). Several statistical tests have been used together to screen and select a group of tracers (the so-called optimum combination of tracers) to establish a statistically verified composite fingerprint based on their discriminatory abilities (e.g., Collins et al., 1997, 1998; Walling et al., 1999; Russell et al., 2001; Gruszowski et al., 2003; Mukundan et al., 2010; Wilkinson et al., 2013; Laceby and Olley, 2014). First, the non-parametric Kruskal–Wallis H-test or Mann–Whitney U-test is used to select tracers that are capable of distinguishing among sediment sources. Second, a bracket test is used to remove non-conservative tracers by comparing tracer concentrations in sediment with those in sources (Mukundan et al., 2010; Collins et al., 2010; Laceby and Olley, 2014). Any tracer that shows higher or lower concentration in sediment than in sources for either mean or individual samples is eliminated. Third, discriminant function analysis (DFA), principal component analysis (PCA), or other multivariate variance analysis have been used to reduce tracer number and to identify an optimum set of tracers (also known as a composite fingerprint) for use in a mixing model (Collins et al., 1997, 1998, 2010; Gruszowski et al., 2003; D'Haen et al., 2012; Gellis and Noe, 2013; Wilkinson et al., 2013). This tracer selection scheme assumes that ability of the tracer in distinguishing among potential sources is fully reflective of its ability in estimating proportional contributions using mixing models, and therefore an optimum composite fingerprint can be selected based on discriminatory ability of the tracer. This assumption is intuitive and has not yet been verified or rigorously tested in the literature.

The linear mixing model, which is based on chemical mass conservation, is the essential basis for providing the relative estimates of sediment source contributions using the fingerprinting technique. If the selected tracers are *conservative*, the mixing model, in principle, should never be tampered by introducing any weighting or correction factors in order to preserve the conservativeness. Any weighting, if not to correct the non-conservativeness, would inadvertently disrupt the conservativeness and potentially bias the contribution estimates, as indicated by the work of Laceby and Olley (2014) and Haddadchi et al. (2014). However, the inclusion of multiple weightings has become central to the contribution estimation, including weightings for particle size correction, organic matter content, within-source variability, and tracer discrimination ability (Collins et al., 1997, 2010; Wallbrink et al., 1998; Wilkinson et al., 2013). The particle size correction can be seen as correcting non-conservativeness of the tracer caused by the enrichment of fines that often have higher tracer concentration. Three different methods were developed in the literature to 'correct' this enrichment effect. First, the fine fractions of <63 μm (known as wash load) were measured in both sources and sediment to circumvent the enrichment issue (e.g., Wallbrink et al., 1998; Walling, 2005; Koiter et al., 2013). Second, a weighting factor based on the ratio of the specific surface areas between a source and sediment was applied in the mixing model (Collins et al., 1997, 1998). Third, precise relationship between particle size composition and tracer concentration for each tracer was developed and used in the correction (Russell et al., 2001). Similarly, organic matter correction factor, which is the ratio of organic contents between sediment and sources, was used (Collins et al., 1997, 1998, 2012). Attention should be paid to the possible overcorrection of organic matter due to the positive correlation between organic matter and particle size distribution (Walling, 2005). The within-source variability and discriminatory ability weightings, giving more weightings in the mixing model to tracers with smaller spatial variations and those with greater discrimination abilities (Collins et al., 1997, 1998, 2010), have received less scrutiny in the sediment tracing literature (Laceby and Olley, 2014). These two weightings, which do not attempt to adjust the non-conservativeness, are questionable, and need to be carefully examined.

In the past decade, more attention has been paid to uncertainty assessment of the estimated source contributions. An analytical approach of the first order approximation was used to develop 95% confidence intervals for the estimated mean proportions of each potential source. Numeric approaches such as Monte Carlo simulation has been increasingly used to simulate probability distributions of estimated source proportions by random or stratified sampling of the probability distributions of each input tracer in each source and/or sediment mixtures (Nagle et al., 2007; Collins et al., 2010, 2012; Stone et al., 2014; Laceby and Olley, 2014). Bayesian uncertainty framework has also been used to simulate probability distributions of estimated source proportions (Small et al., 2002; Koiter et al., 2013; Stewart et al., 2014). Almost all these numeric approaches have been used with only *one set* of the minimum tracers or called the optimum composite fingerprint. As a result, most simulation outputs represent the probability distributions of a source contribution variable rather than the distributions of the contribution *means* of that variable, unless a normal or student distribution is assumed and the *standard error* instead of the standard deviation is used in random sampling to reduce uncertainty associated with the proportion outputs as done by Wilkinson et al. (2013) and Laceby and Olley (2014). However, the assumption of a normal or student distribution is often questionable, as many tracers do not follow such ideal distributions in reality (Olley et al., 2013). Another straight forward way to reduce uncertainty is to simulate the distribution of the estimated *mean* proportions by using as many composite fingerprints as possible. This approach, though running counter to the widely accepted dogma that an optimum composite fingerprint should be used in the source fingerprinting, has never been reported in the literature and deserves a thorough examination.

Overall, significant progresses have been made to advance the fingerprinting technique in the past decades. However, in the past two years, there has been a move to challenge and re-evaluate the underlying assumptions of the fingerprinting technique, methods of data processing, and treatment of uncertainty (Smith et al., 2015). The perception that the fingerprinting technique is a fully functioning research and management tool that delivers accurate estimates of sediment source contributions is misplaced (Smith et al., 2015) and needs to be re-examined. Many fundamental aspects of the technique, including methods of tracer selection, degree of tracer conservativeness, treatments of over-determined mathematical problems, modification of mixing models with weighting and correction factors, structure of objective optimizing functions, and representation of uncertainty, need to be further explored, clarified, and streamlined. Several key studies have been carried out in this regard in the past two years. For example, Laceby et al. (2015) proposed that in addition to the widely used statistical tracer selection, prior knowledge of geochemical, geological, pedogenic, and environmental processes should be considered in tracer selection. Contrary to the widely accepted tracer reduction strategies, Sherriff et al. (2015) reported that uncertainty in source predictions was reduced when more tracers, rather than less, were used in the estimation. They also reported that inclusion of a single non-conservative tracer in the mixing model resulted in large changes in source predictions. The validity of tracer weighting in the mixing models has been questioned by Laceby and Olley (2014) and Haddadchi et al. (2014). Those prominent issues must be judiciously scrutinized in order to further the fingerprinting technique.

The objectives of this study are to (1) evaluate a new approach by using the maximum number of multiple composite fingerprints, as opposed to a single optimum fingerprint as accepted in the literature,