

An extended Bayesian sediment fingerprinting mixing model for the full Bayes treatment of geochemical uncertainties

Richard J. Cooper^{1*} and Tobias Krueger²

¹*School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK; Richard.J.Cooper@uea.ac.uk ; +44(0)1603 592922. *Corresponding author.*

²*IRI THESys, Humboldt University, 10099 Berlin, Germany; tobias.krueger@hu-berlin.de*

Abstract

Recent advances in sediment fingerprinting research have seen Bayesian mixing models being increasingly employed as an effective method to coherently translate component uncertainties into source apportionment results. Here, we advance earlier work by presenting an extended Bayesian mixing model capable of providing a full Bayes treatment of geochemical uncertainties. The performance of the extended full Bayes model was assessed against the equivalent empirical Bayes model and traditional frequentist optimisation. The performance of models coded in different Bayesian software ('JAGS' and 'Stan') was also evaluated, alongside an assessment of model sensitivity to reduced source representativeness and non-conservative fingerprint behaviour. Results revealed comparable accuracy and precision for the full and empirical Bayes models across both synthetic and real sediment geochemistry datasets, demonstrating that the empirical treatment of source data here represents a close approximation of the full Bayes treatment. Contrasts in the performance of models coded in JAGS and Stan revealed that the choice of software employed can impact

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hyp.11154

significantly upon source apportionment results. Bayesian models coded in Stan were the least sensitive to both reduced source representativeness and non-conservative fingerprint behaviour, indicating Stan as the preferred software for future Bayesian sediment fingerprinting studies. Whilst the frequentist optimisation generally yielded comparable accuracy to the Bayesian models, uncertainties around apportionment estimates were substantially greater and the frequentist model was less effective at dealing with non-conservative behaviour. Overall, the effective performance of the extended full Bayes mixing model coded in Stan represents a notable advancement in source apportionment modelling relative to previous approaches. Both the mixing model and the software comparisons presented here should provide useful guidelines for future sediment fingerprinting studies.

Keywords: Tracing; river; apportionment; uncertainty; JAGS; Stan.

1. Introduction

Sediment fingerprinting has become a widely used technique for apportioning the sources of sediment transported through, or deposited within, fluvial environments (Walling, 2013). The technique exploits differences in the geochemistry of potential sediment source areas across a river catchment and contrasts these with the geochemistry of riverine sediments using a mixing model approach (Guzmán et al., 2013). Whilst sediment fingerprinting has been carried out since the 1990s (Collins et al., 1997; Walling et al., 1999), its use has intensified in recent years (Kraushaar et al., 2015; Lamba et al., 2015; Vale et al., 2016) and attention has now shifted onto ways to improve the accuracy and robustness of the procedure (Koiter et al., 2013; Haddadchi et al., 2014; Pulley et al., 2015; Smith et al., 2015). This has included

re-evaluating the criteria for selecting individual fingerprints (Belmount et al., 2014; Laceby et al., 2015; Palazón et al., 2015; Pulley et al., 2016), re-assessing the use of post-hoc data corrections (Smith and Blake, 2014; Laceby and Olley, 2015; Koiter et al., 2015), exploring the effectiveness of land use specific tracers (Blake et al., 2012; Cooper et al., 2015a; Alewell et al., 2016; Reiffarth et al., 2016), and improving methods for accurately quantifying the uncertainties associated with mixing model results (Rowan et al., 2011; Clarke, 2015; Cooper et al., 2015b).

A key advancement in quantifying the uncertainties associated with the sediment fingerprinting procedure has been the emergence of Bayesian mixing models (Massoudieh et al., 2012; D'Haen et al., 2012; Nosrati et al., 2014; Stewart et al., 2015; Abban et al., 2016) as an alternative to the more commonly applied least-squares 'frequentist' mixing model approaches (Walling et al., 2003; Martínez-Carreras et al., 2010; Collins et al., 2013). Uncertainties include spatial and temporal variability in source and riverine sediment geochemistry, analytical instrument error, mixing model error, and unknown residual error such as non-conservative sediment transport (Small et al., 2002; Sherriff et al., 2015). Through the probabilistic treatment of these uncertainties, Bayesian approaches allow for the coherent translation of all sources of analytical and procedural uncertainty into mixing model results within a hierarchical framework, albeit dependent upon error assumptions (Van den Meersche et al., 2008; Cooper et al., 2014a).

Building on previous research in the fields of ecology (Hopkins and Ferguson, 2012; Parnell et al., 2013) and geoscience (Fox and Papanicolaou, 2008), Cooper et al. (2014a) developed a Bayesian mixing model using the open-source software 'JAGS' (Just Another Gibbs Sampler; Plummer, 2003) to apportion the sources of suspended sediment at high-temporal resolution in a lowland river in the United Kingdom. However, JAGS did not allow the full "online" estimation of source and error covariance matrices as it did not ensure positive-

definiteness when sampling these matrices. Hence, a full Bayes treatment of geochemical uncertainties was not possible and consequently an empirical Bayes approach was adopted where some parameters (here the source geochemistry and error covariance matrices) were fixed “offline” at their Maximum Likelihood estimates. Full Bayes approaches, where all model parameters are treated probabilistically, are considered advantageous over empirical approaches as they reduce the likelihood of unrepresentative data, for example where sample numbers are small, biasing mixing model results (Ward et al., 2010). Furthermore, through the incorporation of a separate residual error term, full Bayes mixing models should be able to better account for uncertainties associated with the non-conservative behaviour of fingerprint properties during sediment transport. In frequentist approaches this issue has traditionally been addressed through the application of unrefined particle size and organic matter corrections, the suitability of which has recently been questioned (e.g. Smith and Blake, 2014; Sherriff et al., 2015).

To address existing mixing model deficiencies, we here present an extended version of the Cooper et al. (2014a) model using the alternative open-source software ‘Stan’ (Stan Development Team, 2015), which allows for the full Bayes treatment of all geochemical uncertainties. The primary objectives of this paper are as follows:

- i. To present the structural details of the extended full Bayes sediment fingerprinting mixing model coded in Stan;
- ii. To compare and contrast the accuracy and precision of the full and empirical Bayes mixing models using both synthetic and real sediment geochemistry data;
- iii. To compare the performance of mixing models coded in different software (JAGS vs. Stan);
- iv. To assess mixing model sensitivity to varying levels of representativeness of the source geochemistry data;

- v. To assess mixing model sensitivity to non-conservative fingerprint behaviour;
- vi. To compare Bayesian mixing model performance against the more commonly applied frequentist optimisation approach.

Both the mixing model and software comparisons will provide useful guidelines for fingerprinting studies in other environments and using other models.

2. Methods and Materials

2.1 Mixing model structure

2.1.1 Empirical Bayes

The empirical Bayes sediment fingerprinting model follows that presented by Cooper et al. (2014a) and is summarised by the Directed Acyclic Graph (DAG) in Figure 1a. The model is solved as a mass balance, whereby the concentration of each fingerprint in the target riverine sediment (Y) is obtained from the concentration of each fingerprint in each potential sediment source area (S) multiplied by the proportional sediment contribution (P) derived from that source. This can be summarised by the following likelihood function:

$$(1) \quad L(S, P | Y)$$

An empirical approach is used to estimate the prior distributions for the sediment source area geochemistry (S). Here multivariate normal (MVN) distributions are parameterised using the means (μ^s) and covariance matrices (Σ^s) of the fingerprint concentrations estimated “offline” by Maximum Likelihood from the source area samples. Whilst this has the advantage of reducing model complexity and correlation between parameters, it also reduces model flexibility and can lead to biased estimates where data are unrepresentative.

For the prior probability on the proportions (P), the compositional data are transformed via the isometric log-ratio (ILR) transformation to ensure that all proportions are independent in transformed space and are positive and sum to unity in original space (Egozcue et al., 2003; Parnell et al., 2013). The ILR transformation is specified as:

$$(2) \Phi_i = \text{ILR}(P_i) = V^T \log. \left[\frac{P_{i1}}{g(P_i)}, \dots, \frac{P_{ik}}{g(P_i)} \right]$$

where V is a deterministic $(k-1) \times k$ triangular Helmert matrix used to transform the simplex into orthogonal space and $g(P_i)$ is the geometric mean proportion over all sources $k=1, \dots, K$ for sample i . The actual values of V do not matter here as long as the matrix meets Helmert criteria. Real P values are returned from Φ by exponentiation and re-normalisation (Egozcue et al., 2003):

$$(3) P_i = \text{ILR}^{-1}(\Phi_i)$$

Weakly informative (i.e. relatively flat) normal (N) and inverse gamma (Inv- Γ) distributions are assigned to the Φ prior means (μ^Φ) and variances ($\sigma^{2\Phi}$), respectively (these are so called hyper-parameters). An inverse Wishart distribution (Inv-W) parameterises the combined instrument precision and residual error (Σ^{resZ}). Here, the Wishart scale matrix is defined as the product of an uninformative $J \times J$ identity matrix (I_J) for residual error and an informative covariance matrix (Σ^Z) for instrument error. Σ^Z , like the source covariance matrices above, was derived empirically by Maximum Likelihood from 42 repeat analyses of a sediment standard. This instrument error (Σ^Z) is subtracted offline from the total variability in the source fingerprint data (Σ^{SZ}) to isolate the spatial variability in the source data (Σ^S) used for estimating the source distributions (S).

See Cooper et al. (2014a) for a sensitivity analysis of the model with respect to the various error assumptions made here. Table S1 provides definitions for the mixing model components, whilst the complete posterior distribution of the empirical Bayes model is summarised as follows:

(4)

$$\begin{aligned}
 & p(\mu, S, P, \Phi, \mu^\Phi, \sigma^{2\Phi}, \Sigma^{resZ} | Y) \\
 & \propto p(Y | \mu, \Sigma^{resZ}) \times p(S | \mu^S, \Sigma^S) \times p(\Phi | \mu^\Phi, \sigma^{2\Phi}) \times p(\Sigma^{resZ}) \times p(\mu^\Phi) \\
 & \times p(\sigma^{2\Phi})
 \end{aligned}$$

2.1.2 Extended full Bayes

The full Bayes mixing model (Figure 1b) extends the empirical model by treating the hyper-parameters of both the source geochemistry distributions and the instrument error as random variables with their own prior distributions. Additional data is then supplied to estimate these priors directly through the MCMC algorithm, just as is done with the Y data. As such, all priors are integrated out probabilistically during the numerical solution rather than some of them being fixed “offline”. This is true to the Bayesian paradigm and should, theoretically, increase model flexibility and reduce the risk of model bias from unrepresentative data. Specifically, it relaxes the assumption that the geochemistry of the sampled source area sediments is fully representative of the entire catchment wide geochemical variability.

The prior means of the sources (μ^S) were assigned uninformative (i.e. flat) normal distributions, $N(10,10)$, based on the fingerprint data being supplied as percentage concentrations on a scale of 0–100% (see section 2.5). Uninformative inverse Wishart distributions $\text{Inv-W}(I_J, J)$ were assigned to the source covariance matrices (Σ^S). In turn, the

posteriors of μ^S and Σ^S were estimated online based on the actual observed source data (Y^S). Instrument precision (Σ^Z) and residual error (Σ^{res}) are assigned separate inverse Wishart distributions, as opposed to the combined Σ^{resZ} of the empirical model, with the posterior of Σ^Z being estimated from the observed, normalised, instrument error data (Y^Z). In total, 10 nodes of random variables were estimated by MCMC sampling in the full Bayes model compared with 7 nodes in the empirical version. The extended full Bayes model is summarised as follows:

(5)

$$\begin{aligned}
 & p(\mu, S, \mu^S, \Sigma^S, P, \Phi, \mu^\Phi, \sigma^{2\Phi}, \Sigma^{res}, \Sigma^Z | Y, Y^S, Y^Z) \\
 & \propto p(Y | \mu, \Sigma^{res}, \Sigma^Z) \times p(S | \mu^S, \Sigma^S) \times p(Y^S | \mu^S, \Sigma^S, \Sigma^Z) \times p(\Phi | \mu^\Phi, \sigma^{2\Phi}) \\
 & \times p(Y^Z | \Sigma^Z) \times p(\Sigma^{res}) \times p(\Sigma^Z) \times p(\mu^S) \times p(\Sigma^S) \times p(\mu^\Phi) \times p(\sigma^{2\Phi})
 \end{aligned}$$

2.2 Bayesian software: JAGS and Stan

Two types of open source software for running probabilistic Bayesian hierarchical models with Markov Chain Monte Carlo (MCMC) sampling are ‘JAGS’ (‘Just another Gibbs sampler’; Plummer, 2003) and ‘Stan’ (Stan Development Team, 2015) – both of which are interfaced within the R environment (R Core Team, 2016). Although either can be used for developing sediment fingerprinting mixing models, there are important implementation differences between the software which affect how the models perform. Firstly, Stan and JAGS use different MCMC sampling algorithms to estimate the posterior distribution. Whilst JAGS uses Gibbs sampling (Gelfand, 2000), Stan uses a variant of Hamiltonian Monte Carlo (HMC) known as a ‘No-U-Turn’ sampler. This No-U-Turn sampler avoids the random walk behaviour and sensitivity to correlated parameters common to other MCMC samplers, thus enabling the model to more quickly converge on the optimal solution (Hoffman and Gelman,

2014). These contrasting MCMC procedures result in differing explorations of the model parameter space, increasing the possibility of the two procedures yielding different apportionment results when samplers do not converge. Additionally, there are significant differences in the handling of covariances, with ordinary covariance matrices implemented in Stan, whilst inverse covariance (i.e. precision) matrices have to be parameterised in JAGS. This can lead to sizable differences in covariance values between model versions when matrices are inverted offline (empirically) by approximation and are forced into the nearest positive-definite matrix before being supplied to the mixing model – an essential procedural step in the JAGS operation. Lastly, Stan allows a greater range of algebraic operations to be performed compared to JAGS, thus facilitating enhanced model development.

2.3 Frequentist optimisation

For comparison of these Bayesian models with the frequentist optimisation approaches adopted in the majority of previous sediment fingerprinting studies (e.g. Walling, 2013), a least-squares regression model based on Collins et al. (1997) was formulated. The frequentist model optimisation was solved for P by minimising the sum of squared residuals (SSR):

(6)

$$\text{SSR} = \sum_{j=1}^J \left(Y_j - \sum_{k=1}^K S_{j,k} P_k \right)^2$$

whilst satisfying the following constraints:

(7)

$$\sum_{k=1}^K P_k = 1$$

and

(8)

$$P_k \geq 0$$

Similar to recent practice (e.g. Collins et al., 2012), instrument error and variability in source (S) and target (Y) fingerprint geochemistry was incorporated into the model by nesting the Maximum Likelihood optimisation step within an ordinary Monte Carlo iteration. This involved sampling from multivariate normal distributions for both the S and Y parameters using the empirical source mean (μ^s), source covariance (Σ^s) and instrument error (Σ^Z) estimates.

2.4 Mixing model versions

Three versions of the Bayesian mixing model were formulated (Table 1): model 1 (M1) was the empirical Bayes model of Cooper et al. (2014a) run in JAGS; model 2 (M2) was the same empirical Bayes model run in Stan; and model 3 (M3) was the extended full Bayes model run in Stan. Model 4 (M4) was the frequentist approach. All four models were run in the R environment. M1 was run for 750,000 iterations, with a 250,000 sample burn-in to reduce the impact of the initial starting conditions, and a jump length of 225. The jump length determines how far the sampler moves within the parameter space between each step of the MCMC walk. In setting a longer jump length exploration of the entire parameter space occurs faster and autocorrelation between runs is minimised by reducing the number of subsequent sample draws from the same region. However, a balance has to be found as setting the jump length too long can result in a low acceptance rate, meaning a greater number of iterations are needed to achieve convergence of the posterior distribution.

The higher efficiency of the No-U-Turn sampler allowed a reduced number of iterations (10,000), burn-in (5,000) and jump length (25) for Stan models M2 and M3. For all Bayesian models, three MCMC chains were run in parallel from random initial conditions and trace plots of the parameter distributions were inspected for mixing. Models were run until convergence was achieved for all variables, with convergence diagnostics performed via the ‘coda’ package within R (Plummer et al., 2006). The frequentist model optimisation was run for 10,000 iterations using the ‘limSolve’ R package (Van den Meersche et al., 2008).

2.5 Data collection

2.5.1 Characterising sediment sources

Sediment source area geochemistry was characterised using real data collected as part of the River Wensum Demonstration Test Catchment (DTC) project, of which comprehensive details can be found in Cooper et al. (2015b). To summarise, four potential sediment source areas were identified across a 5.4 km² section of the lowland, arable, Blackwater sub-catchment of the River Wensum, Norfolk, UK (52°47'14"N, 1°07'42"E). These were eroding stream channel banks, subsurface agricultural field drains, arable topsoil, and damaged road verges (Figure 2). From each source area, 30 soil/sediment samples were collected, wet sieved to <63 µm to extract the fine clay-silt fraction, and transferred onto quartz fibre filter papers. These filter papers were then analysed by X-ray fluorescence spectroscopy (XRFS) to determine the geochemistry (wt. %) following the method of Cooper et al. (2014b) which had previously been successfully applied in this catchment. In total, concentrations of eight major elements (Al, Ca, Ce, Fe, K, Mg, Na and Ti) were determined and selected as fingerprints for use in the mixing models. Prior to running the models, the geometry of the source geochemistry mixing space was examined via a principal component analysis (Figure 3). The

fingerprint loadings (right-hand panels) revealed calcium (Ca) to be the most powerful discriminator of the surface (road verge and topsoil) and subsurface (channel bank and field drain) sources, whilst iron (Fe) provided strong discrimination of road verge versus topsoil and channel bank versus field drain sources. However, significant overlap in the geochemical ranges of both channel bank and field drain sediments made differentiation difficult and the geochemical data for these two sources were consequently merged into a single combined 'subsurface' sediment source prior to running the models.

2.5.2 Reducing source representativeness

As described previously, one of the main advantages of adopting a full Bayes approach is to minimise the risk of unrepresentative empirical data biasing mixing model results. This is particularly true where the number of samples collected from each source area is small, with Ward et al. (2010) cautioning that full Bayes approaches should be adopted if there are <20 samples per source. Here we assess the impact of reduced source representativeness on model accuracy and precision by re-running each model four times for the six synthetic mixtures with source distributions parameterised using a decreasing number of sediment source samples (Table 2).

2.5.3 Characterising target river sediment mixtures

For the target river sediment (Y), both real and synthetic data were used. Firstly, a set of six synthetic mixtures were generated within the R environment to incorporate a wide range of subsurface (0–75%), road verge (0–60%) and topsoil (12.5–100%) contributions (Table 3). The mean concentrations for the eight elemental fingerprints in each source area were

considered to represent 100% contribution from that particular source (e.g. mixture 1 in Table 3). Derivative mixtures of this were then obtained by multiplying the mean concentrations by the relative proportions of each source (e.g. 0.5 x road verge mean + 0.5 x topsoil mean for mixture 2). Using synthetic mixtures with known source contributions enables a robust assessment of mixing model accuracy and precision.

Then, to assess model performance using real target data, instream suspended particulate matter (SPM) samples were collected at the outlet of the 5.4 km² portion of the Blackwater sub-catchment during a heavy rainfall event in February 2013. A bankside ISCO automatic sampler (Teledyne ISCO, Lincoln, NE) was remotely activated to collect a 1 L stream water sample every two hours over a 48 hour period. Use of automatic samplers can introduce sampling bias due to the preferential uptake of finer SPM in the suction line (Krueger et al., 2009), but this issue was largely circumvented here by wet sieving the resulting 24 samples to <63 µm on returned to the laboratory. The samples were then vacuum filtered onto quartz fibre filter papers to extract the SPM and analysed by XRF to determine the geochemistry (wt. %) of the target riverine sediment.

2.5.4 Non-conservative target mixtures

To assess the ability of each mixing model to handle the non-conservative behaviour of fingerprint properties during sediment transport, each model was re-run for the six synthetic mixtures using four target geochemistry datasets that had been deliberately manipulated/corrupted offline to simulate downstream enrichment. This involved increasing the concentrations of all eight target sediment fingerprints by 10, 20, 50 and 80% for each of the six target mixtures (Table S2). Whilst we acknowledge that in a real-world situation all fingerprints would not necessarily be enriched equally during downstream transport, the data corruption tested here provides a 'worst case scenario' and an ultimate test for the models in

which none of the fingerprints in the target sediment have maintained their original composition. The average deviation between the model estimated proportions and the actual true proportions could then be used to infer model sensitivity to non-conservative fingerprint behaviour.

3. Results and Discussion

3.1 Model comparisons: synthetic data

Sediment source apportionment results for the six synthetic target mixtures are presented in Figure 4 and summarised in Table 4. These results reveal fairly consistent apportionment estimates both between different model software (M1 and M2) and between full and empirical Bayes approaches (M2 and M3). The average deviation across the six mixtures between the median apportionment values estimated by the model and the actual true proportions (i.e. the model accuracy) was 12.9% (range = 0.4–34.3%) for the empirical JAGS model, 11.6% (range = 0.5–27.4%) for the empirical Stan model and 11.4% (range = 0.6–26.3%) for the full Bayes Stan model. The frequentist model had slightly higher accuracy, with a mean deviation from the true contribution of 8.9% (range = 0.0–27.8%). The greatest Bayesian model accuracy was observed for mixture 4 (33.3% subsurface, 33.3% road verge and 33.3% topsoil), with an average deviation from the actual contribution of just 5.9% observed across all three sources. This can be explained by the mean hyper-parameter distribution on the transformed proportions (μ^ϕ) being a normal distribution centred on zero ($N(0, 1)$), which, after retransformation, equates to equal contributions (i.e. 33.3%) from all sources. The lowest accuracy was recorded for mixture 5 (50% subsurface and 50% topsoil) where all models overestimated road verge contribution and underestimated topsoil and

subsurface proportions, resulting in an average deviation across the three sources of 20.7% from the actual contributions.

In terms of the estimated apportionment uncertainty, mean 95% credible interval widths varied from 55.3% (range = 19.6–82.3%) for the empirical Stan model, to 56.4% (range = 21.8–80.1%) for the full Bayes Stan model and 61.4% (range = 30.8–90.4%) for the empirical JAGS model. The frequentist model M4 had the highest degree of uncertainty with a mean confidence interval width of 70.0% (range = 12.7–100.0%) across the six mixtures. Estimates for mixture 2 yielded the greatest amount of uncertainty for an individual mixture amongst the Bayesian models, with a mean credible interval width of 65.6%. Estimated proportions for mixture 1 had the lowest uncertainty, with a mean credible interval width of 47.8%.

Overall, all models showed the highest level of accuracy and precision when estimating the contribution of sediment derived from the subsurface source. This reflects the relative uniqueness of the subsurface source geochemistry compared with topsoil and road verge sources whose geochemical ranges were more closely aligned (Figure 3). This characteristic of the source dataset reduces the ability of the mixing model to successfully differentiate between road verge and topsoil sources, thus resulting in reduced model accuracy and increased uncertainty. The Stan models (M2 and M3) did, however exhibit marginally improved discrimination between road verge and topsoil sources compared to the JAGS model, with both higher accuracy and narrower credible interval widths observed for these two sources. This higher uncertainty of the JAGS model was most apparent for mixtures 1 (100% topsoil mix) and 2 (50% road verge, 50% topsoil) where the actual subsurface contribution was zero and hence the target geochemistry sat directly within the road verge–topsoil mixing space. This seems to suggest that where geochemical distributions of the

source samples overlap, the Stan models may be able to achieve better source group differentiation than the JAGS model under the conditions of this study.

3.2 Model comparisons: real data

Source apportionment results for SPM transported in the River Blackwater during a heavy precipitation event in February 2013 are presented in Figure 5, whilst Figure S1 presents time-series of changing SPM geochemistry. During the event, 12.8 mm of rainfall fell over a period of 10 h resulting in a five-fold rise in river discharge (0.03 to $0.15 \text{ m}^3 \text{ s}^{-1}$) and a rapid increase in SPM concentration (9 to 176 mg L^{-1}) within 2 h of peak rainfall intensity. Prior to the event, all three Bayesian models estimated similarly high subsurface (53–54%), low topsoil (11–13%) and intermediate road verge (29–33%) median contributions. This differed from the frequentist model which estimated a substantially higher median road verge contribution (62%) and lower subsurface (38%) and topsoil (0%) proportions. As SPM concentrations increased during the onset of the heaviest rainfall ~8 h into the monitoring period, all models estimated a rapid and pronounced increase in topsoil contribution and corresponding declines in subsurface and, to a lesser extent, road verge material. Median topsoil apportionment estimates increased to 47% (4–85% at the 95% credible interval) for the empirical JAGS model, 66% (13–93%) for the empirical Stan model and 65% (14–94%) for the full Bayes Stan model. The frequentist model estimated the highest median topsoil contribution of 90% (33–100% at the 95% confidence interval) during this period.

As the event progressed over the following 24–36 h, SPM concentrations gradually reduced back down to pre-event levels (12 mg L^{-1}) and estimated subsurface contributions increased consistently across the three Bayesian models to once again become the dominant sediment source (median = 43–51%; 95% credible interval = 28–80%). The frequentist model also

estimated an increase in subsurface sediment contribution by the end of the monitoring period (median = 53%), but with greater uncertainty (95% confidence interval = 0–100%). In fact, as with the synthetic mixtures, the frequentist model yielded the most uncertain source apportionment estimates for all three sources across the whole precipitation event, with mean 95% confidence interval widths of 79% for the subsurface, 95% for road verges and 53% for topsoil. Conversely, the full and empirical Bayes Stan models yielded more precise apportionment estimates, with near identical mean 95% credible interval widths of 49% for the subsurface, 61% for road verges and 59% for topsoil for both models.

Differences in apportionment estimates between the four model versions consequently impacted upon the estimated SPM load transported during the 48 h event (Table 5). Very similar SPM loads were estimated for the empirical and full Bayes Stan models, with 604–608 kg derived from topsoil, 427–428 kg from the subsurface and 351–357 kg from road verges based on the median estimate. This equates to 43% (6–77% at the 95% credible interval) of the ~1390 kg of SPM transported to originate from topsoil, 31% (13–63%) from the subsurface and 26% (4–69%) from road verges. Conversely, the empirical JAGS model estimated a substantially greater load of SPM from road verges (562 kg) and lower loads from topsoil (445 kg) and the subsurface (386 kg). These JAGS estimates translate into 40% (9–78% at the 95% credible interval) of total SPM transport originating from road verges, 32% (2–70%) from topsoil and 28% from the subsurface (13–57%). The frequentist model estimated the highest topsoil (650 kg) and lowest subsurface (326 kg) loads of the four model versions. This equates to 47% (13–79% at 95% confidence interval) of SPM originating from topsoil, 29% (0–91%) from road verges and 24% (0–62%) from the subsurface.

3.3 Model sensitivity to reduced source representativeness

The impact on mixing model accuracy and uncertainty of reducing the representativeness of the source area geochemistry data is shown in Figure 6. As the number of source area samples was reduced from 120 to 80, there was no obvious deterioration in the accuracy of any of the four model versions (Figure 6A). However, further reductions in the number of source samples down to 40 and then 15, resulted in a pronounced contrast in the accuracy of apportionment results between models. The performance of the empirical JAGS model deteriorated most severely, with the average deviation between the median estimated apportionment values and the actual true proportions increasing significantly ($p < 0.05$) from 12.9% at 120 samples to 21.5% at 15 samples. The Stan empirical model deviation increased from 11.6% at 120 samples to 14.1% at 15 samples, whilst the frequentist model deviation increased from 8.9% to 12.1%. The full Bayes Stan model was the least sensitive to changes in the number of source samples, with a non-significant ($p > 0.05$) increase in mean deviation of just 0.8% from 11.4% at 120 samples to 12.2% at 15 samples. This is consistent with the observation of Ward et al. (2010) that full Bayesian approaches are advantageous where sample numbers are small.

There was an inconsistent response in model uncertainty to reducing sample numbers (Figure 6B). The average 95% confidence interval widths of the frequentist model demonstrated the greatest sensitivity, increasing from 70.0% at 120 samples to 80.1% at 15 samples, albeit non-significantly ($p > 0.05$) due to considerable scatter. Conversely, the JAGS model showed a small, yet consistent, decline in the mean 95% credible interval width from 61.4% at 120 samples to 56.7% with 15 samples. The empirical and full Bayes Stan models were the least sensitive to changing source sample numbers, with no significant ($p > 0.05$) change in credible interval width between 120 and 15 samples.

The reduced sensitivity of the full Bayes version to source geochemistry representativeness arises because assigning uninformative prior distributions to the hyper-parameters of the sources relaxes the assumption that the geochemistry of the source samples is fully representative of the entire catchment wide geochemical variability. Consequently, the model is more flexible and less likely to be biased by unrepresentative data. In this respect, it is slightly surprising that the performance of the empirical Stan model did not deteriorate further as the sample numbers were reduced. This suggests that for the dataset used here, the empirical treatment of the source data does represent a good approximation of the full Bayes treatment and that either approach could reliably be adopted.

However, a key reason behind the higher sensitivity of the empirical JAGS model relates to the parameterisation of precision matrices instead of the covariance matrices employed in Stan. As the number of source samples is reduced, correlation between the source geochemical fingerprints is altered, thus altering the covariance structure of the source dataset. When the covariance matrices are inverted offline (empirically) to form precision matrices by approximation, which necessitates forcing the precision matrices into the nearest positive-definite matrices before input to JAGS, the precision matrix estimates can become biased resulting in substantially different prior source distributions (Figure 7). Here, this leads to a decrease in the accuracy and uncertainty of the JAGS model results as the number of source samples is reduced.

3.4 Model sensitivity to non-conservative behaviour

The impact on mixing model performance of corrupting the target geochemistry to represent non-conservative fingerprint behaviour can be seen in Figure 8. The empirical JAGS model was the most sensitive to the corruption of the target dataset, with the average deviation between the median estimated apportionment values and the actual proportions increasing significantly ($p < 0.05$) from 12.9% with no corruption to 33.6% at 80% corruption (Figure 8A). The frequentist model also displayed relatively high sensitivity to tracer corruption, with mean deviation increasing significantly ($p < 0.05$) from 8.9% with the uncorrupted target dataset to 21.5% when target geochemistry values were increased by 80%. These observations are similar to that previously reported by Sherriff et al. (2015). Conversely, both empirical and full Bayes Stan models exhibited minimal deterioration in model accuracy as target fingerprint corruption increased, with a non-significant ($p > 0.05$) increase in deviation of <6% at 80% corruption. This clearly demonstrates that the Stan models are able to cope better with non-conservative fingerprint behaviour than either the JAGS or frequentist approaches. The reduced performance of the JAGS model with increased fingerprint corruption is again an artefact of the biased covariance structures that are produced when inverting to precision matrices offline (Figure 7).

As observed when reducing source representativeness, there was an inconsistent response in mixing model uncertainty to the corruption on the target geochemistry (Figure 8B). The 95% credible interval widths of the JAGS model declined significantly ($p < 0.05$) from 61.4% to 39.6% as corruption increased to 80%. This, together with the decreased accuracy, arose because as the degree of corruption increased the model increasingly estimated contributions almost entirely of subsurface origin. Whilst the accuracy of both Stan models was largely insensitive to non-conservative behaviour, model uncertainty did increase by ~5% between 0% and 80% fingerprint corruption, albeit with considerable scatter and no significant trend

($p > 0.05$). Frequentist model uncertainty remained relatively unchanged, with a 3% decline in confidence interval width at 80% corruption.

3.5 Significance and recommendations

For both artificial and real sediment mixtures, the results presented here have demonstrated a clear consistency in source apportionment estimates between mixing models adopting either an empirical or a full Bayes treatment of geochemical uncertainties. Therefore, for the dataset used here, we can say that the empirical treatment of the source data does represent a good approximation of the full Bayes treatment and that either method would be suitable under these conditions. Whilst empirical Bayes might be preferred for numerical efficiency (although efficiency gains are small in Stan), we recommend that future studies stay true to the fundamental principles of Bayesian statistics and adopt the full Bayesian approach to minimise the risk of a more unrepresentative dataset than that utilised here biasing the apportionment results.

Contrasts in the performance of mixing models coded in JAGS and Stan reveal that the choice of mixing model software employed can impact significantly upon source apportionment results. This is most apparent in the increased sensitivity of the JAGS model to both reduced source representativeness and non-conservative tracer behaviour. JAGS necessitates the use of inverse covariance (i.e. precision) matrices to characterise the uncertainty around source, target and instrument error, and it is the bias that can be introduced during the “offline” estimation of these precision matrices that results in the reduced performance of the JAGS model. As shown in Figure 7, distributions parameterised by covariance matrices more closely approximate the actual empirical data. Whilst further research comparing the relative performance of JAGS and Stan coded mixing models is

required to confirm the results presented here, these initial findings indicate Stan as the preferred software for future Bayesian sediment fingerprinting studies.

Comparison of Bayesian and frequentist mixing models reveals generally comparable accuracy in source apportionment estimates but with substantially greater uncertainty generated by the frequentist approach. As was previously demonstrated by Cooper et al. (2014a), this is in large part due to the bimodality of frequentist source apportionment results in which 0% and 100% are often the most commonly estimated source contributions. This, arguably unrealistic behaviour, arises from the linear optimisation forcing a fit onto single draws from the source and target datasets. The effect is avoided in Bayesian inference where the entire distributions of all parameters are explored simultaneously resulting in more realistic posterior distributions. The frequentist model is also less effective at dealing with non-conservative tracer behaviour than either of the Stan Bayesian models. These results, combined with the fact that frequentist optimisation lacks the structural flexibility to coherently translate all sources of uncertainty into mixing model results, means that adopting Bayesian inference in future sediment fingerprinting studies is preferable to the more commonly applied frequentist approach.

4. Conclusions

The empirical Bayes mixing model published by Cooper et al. (2014a) represented a notable improvement in the handling of uncertainties associated with the sediment fingerprinting technique compared with commonly employed least-squares frequentist optimisation routines. However, in fixing the prior parameters of the source geochemistry offline, the model did not stay true to the Bayesian paradigm of treating all model parameters probabilistically, thus increasing the risk of the model results being biased by

unrepresentative input data. The extended full Bayes model presented here addresses this deficiency. Whilst it is recommended that this extended full Bayes model supersedes the earlier empirical version, the empirical model coded in Stan remains a satisfactory approximation. In all cases, the frequentist optimisation generated substantially greater uncertainty around source apportionment estimates relative to the Bayesian models and was less effective at dealing with non-conservative tracer behaviour. Differences in the performance of models coded in JAGS and Stan, specifically the lower sensitivity of the Stan models to both reduced source representativeness and non-conservative fingerprint behaviour, provisionally highlight Stan as the preferred software for future fingerprinting studies. Further investigation into the comparative performance of different Bayesian mixing model software is highly recommended as the research community looks to continue advancing existing sediment fingerprinting techniques.

Acknowledgements

TK acknowledges funding from the German Excellence Initiative through IRI THESys. We are grateful to the editor and two anonymous reviewers whose constructive comments helped improve an earlier version of the manuscript.

Supporting Information

Table S1: Summary definitions for the mixing model components.

Table S2: Geochemical fingerprints for non-conservative transport.

Figure S1: Time-series of River Wensum SPM geochemistry during February 2013.

Code: Stan model code for running the empirical and extended full Bayes models.

References

- Abban B, Papanicolaou AN, Cowles MK, Wilson CG, Abaci O, Wacha K, Schilling K, Schnoebelen D. 2016. An enhanced Bayesian fingerprinting framework for studying sediment source dynamics in intensively managed landscapes. *Water Resources Research*, 52. DOI: 10.1002/2015WR018030.
- Alewell C, Birkholz A, Meusburger K, Schindler Wildhaber Y, Mabit L. 2016. Quantitative sediment source attribution with compound-specific isotope analysis in a C3 plant-dominated catchment (central Switzerland). *Biogeosciences*, 13, 1587-1596. DOI: 10.5194/bg-13-1587-2016.
- Belmont P, Willenbring JK, Schottler SP, Marquard J, Kumarasamy K, Hemmis JM. 2014. Toward generalizable sediment fingerprinting with tracers that are conservative and nonconservative over sediment routing timescales. *Journal of Soils and Sediments*, 14, 1479-1492. DOI: 10.1007/s11368-014-0913-5.
- Blake WH, Ficken KJ, Taylor P, Russell MA, Walling DE. 2012. Tracing crop-specific sediment sources in agricultural catchments. *Geomorphology*, 139-140, 322-329. DOI:10.1016/j.geomorph.2011.10.036.
- Clarke RT. 2015. A bootstrap calculation of confidence regions for proportions of sediment contributed by different source areas in a 'fingerprinting' model. *Hydrological Processes*, 29, 2679-2703. DOI: 10.1002/hyp.10397.
- Collins AL, Walling DE, Leeks GJL. 1997. Source type ascription for fluvial suspended sediment based on a quantitative composite fingerprinting technique. *Catena*, 29, 1-27. DOI:10.1016/S0341-8162(96)00064-1.

- Collins AL, Zhang Y, Walling DE, Grenfell SE, Smith P, Grischeff J, Locke A, Sweetapple A, Brogden D. 2012. Quantifying fine-grained sediment sources in the River Axe catchment, southwest England: application of a Monte Carlo numerical modelling framework incorporating local and genetic algorithm optimisation. *Hydrological Processes*, 26, 1962-1983. DOI: 10.1002/hyp.8283.
- Collins AL, Zhang YS, Hickinbotham R, Bailey G, Darlington S, Grenfell SE, Evans R, Blackwell M. 2013. Contemporary fine-grained bed sediment sources across the River Wensum Demonstration Test Catchment, UK. *Hydrological Processes*, 27, 857-884. DOI: 10.1002/hyp.9654.
- Cooper RJ, Krueger T, Hiscock KM, Rawlins BG. 2014a. Sensitivity of fluvial sediment source apportionment to mixing model assumptions: A Bayesian model comparison. *Water Resources Research* 50: 9031-9047. DOI: 10.1002/2014WR016194.
- Cooper RJ, Rawlins BG, Lézé B, Krueger T, Hiscock K. 2014b. Combining two filter paper-based analytical methods to monitor temporal variations in the geochemical properties of fluvial suspended particulate matter. *Hydrological Processes* 28: 4042-4056. DOI: 10.1002/hyp.9945.
- Cooper RJ, Pedentchouk N, Hiscock KM, Disdle P, Krueger T, Rawlins BG. 2015a. Apportioning sources of organic matter in streambed sediments: An integrated hydrogen and carbon stable isotope approach. *Science of the Total Environment*, 520, 187-197. DOI: 10.1016/j.scitotenv.2015.03.058.
- Cooper RJ, Krueger T, Hiscock KM, Rawlins BG. 2015b. High-temporal resolution fluvial sediment source fingerprinting with uncertainty: a Bayesian approach, *Earth Surface Processes and Landforms* 40: 78-92. DOI: 10.1002/esp.3621.
- D'Haen K, Verstraeten G, Duser B, Degryse P, Haex J, Waelkens M. 2012. Unravelling changing sediment sources in a Mediterranean mountain catchment: a Bayesian fingerprinting approach. *Hydrological Processes*, 27, 896-910. DOI: 10.1002/hyp.9399.

- Egozcue J, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. 2003. Isometric log ratio transformations for compositional data analysis. *Mathematical Geology*, 35, 279-300. DOI: 10.1023/A:1023818214614.
- Fox JF, Papanicolaou AN. 2008. An un-mixing model to study watershed erosion processes. *Advances in Water Resources*, 31, 96-108. DOI:10.1016/j.advwatres.2007.06.008.
- Gelfand AE. 2000. Gibbs sampling. *Journal of the American Statistical Association*, 95, 1300-1304. DOI: 10.1080/01621459.2000.10474335.
- Guzmán G, Quinton JN, Nearing MA, Mabit L, Gómez JA. 2013. Sediment tracers in water erosion studies: current approaches and challenges. *Journal of Soils and Sediments*, 13, 816-833. DOI: 10.1007/s11368-013-0659-5.
- Haddadchi A, Olley J, Laceby P. 2014. Accuracy of mixing models in predicting sediment source contributions. *Science of the Total Environment*, 497-498, 139-152. DOI: 10.1016/j.scitotenv.2014.07.105.
- Hoffman MD, Gelman A. 2014. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 1593-1623.
- Hopkins JB, Ferguson JM. 2012. Estimating the diets of animals using stable isotopes and a comprehensive Bayesian mixing model. *PLoS ONE*, 7, e28478. DOI:10.1371/journal.pone.0028478.
- Koiter AJ, Owens PN, Petticrew EL, Lobb DA. 2013. The behavioural characteristics of sediment properties and their implications for sediment fingerprinting as an approach for identifying sediment sources in river basins. *Earth-Science Reviews* 125, 24-42. DOI: 10.1016/j.earscirev.2013.05.009.
- Koiter AJ, Owens PN, Petticrew EL, Lobb DA. 2015. The role of gravel channel beds on the particle size and organic matter selectivity of transported fine-grained sediment: implications for sediment

fingerprinting and biogeochemical flux research. *Journal of Soils and Sediments*, 15, 2174-2188.

DOI: 10.1007/s11368-015-1203-6.

Kraushaar S, Schumann T, Ollesch G, Schubert M, Vogel H-J, Siebert C. 2015. Sediment fingerprinting in northern Jordan: element-specific correction factors in a carbonatic setting. *Journal of Soils and Sediments*, 15, 2155-2173. DOI: 10.1007/s11368-015-1179-2.

Krueger T, Quinton JN, Freer J, Macleod CJA, Bilotta GS, Brazier RE, Butler P, Haygarth PM. 2009. Uncertainties in data and models to describe event dynamics of agricultural sediment and phosphorus transfer. *Journal of Environmental Quality*, 38, 1137-1148. DOI: 10.2134/jeq2008.0179.

Lacey JP, Olley J. 2015. An examination of geochemical modelling approaches to tracing sediment sources incorporating distribution mixing and elemental correlations. *Hydrological Processes*, 29, 1669-1685. DOI: 10.1002/hyp.10287.

Lacey JP, McMahon J, Evrard O, Olley J. 2015. A comparison of geological and statistical approaches to element selection for sediment fingerprinting. *Journal of Soils and Sediments*, 15, 2117-2131. DOI: 10.1007/s11368-015-1111-9.

Lamba J, Karthikeyan KG, Thompson AM. 2015. Apportionment of suspended sediment sources in an agricultural watershed using sediment fingerprinting. *Geoderma*, 239-240, 25-33. DOI: 10.1016/j.geoderma.2014.09.024.

Martínez-Carreras N, Krein A, Udelhoven T, Gallart F, Iffly JF, Hoffmann L, Pfister L, Walling DE. 2010. A rapid spectral-reflectance-based fingerprinting approach for documenting suspended sediment sources during storm runoff events. *Journal of Soils and Sediments*, 10, 400-413. DOI: 10.1007/s11368-009-0162-1.

Massoudieh A, Gellis A, Banks WS, Wiczorek ME. 2012. Suspended sediment source apportionment in Chesapeake Bay watershed using Bayesian chemical mass balance receptor modelling. *Hydrological Processes*, 27, 3363-3374. DOI: 10.1002/hyp.9429.

- Nosrati K, Govers G, Semmens BX, Ward EJ. 2014. A mixing model to incorporate uncertainty in sediment fingerprinting. *Geoderma*, 217-218, 173-180. DOI: 10.1016/j.geoderma.2013.12.002.
- Palazón L, Latorre B, Gaspar L, Blake WH, Smith HG, Navas A. 2015. Comparing catchment sediment fingerprinting procedures using an auto-evaluation approach with virtual sample mixtures. *Science of the Total Environment*, 532, 456-466. DOI: 10.1016/j.scitotenv.2015.05.003.
- Parnell AC, Phillips DL, Bearhop S, Semmens BX, Ward EJ, Moore JW, Jackson AL, Inger R. 2013. Bayesian stable isotope mixing models. *Environmetrics*, 24, 387-399. DOI: 10.1002/env.2221.
- Plummer M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, Proceeding of the 3rd international workshop on distributed statistical computing, Vienna, Austria.
- Plummer M, Best N, Cowles K, Vines K. 2006. R Package 'coda': convergence diagnosis and output analysis for MCMC.
- Pulley S, Foster I, Antunes P. 2015. The uncertainties associated with sediment fingerprinting suspended and recently deposited fluvial sediment in the Nene river basin. *Geomorphology*, 228, 303-319. DOI: 10.1016/j.geomorph.2014.09.016.
- Pulley S, Foster I, Collins AL. 2016. The impact of catchment source group classification on the accuracy of sediment fingerprinting outputs. *Journal of Environmental Management*, 1-11. DOI: 10.1016/j.jenvman.2016.04.048.
- Reiffarth DG, Petticrew EL, Owens PN, Lobb DA. 2016. Sources of variability in fatty acid (FA) biomarkers in the application of compound-specific stable isotopes (CSSIs) to soil and sediment fingerprinting and tracing: A review. *Science of the Total Environment*, 565, 8-27. DOI: 10.1016/j.scitotenv.2016.04.137.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

- Rowan JS, Black S, Franks SW. 2011. Sediment fingerprinting as an environmental forensics tool explaining cyanobacteria blooms in lakes. *Applied Geography*, 32, 832-843. DOI:10.1016/j.apgeog.2011.07.004.
- Sherriff SC, Franks SW, Rowan JS, Fenton O, Ó'hUallacháin D. 2015. Uncertainty-based assessment of tracer selection, tracer non-conservativeness and multiple solutions in sediment fingerprinting using synthetic and field data. *Journal of Soils and Sediments*, 15, 2101-2116. DOI: 10.1007/s11368-015-1123-5.
- Small IF, Rowan JS, Franks SW. 2002. Quantitative sediment fingerprinting using a Bayesian uncertainty estimation framework, in *The structure function and management implications of fluvial sedimentary systems*, edited by F. J. Dyer, M.C. Thomas, and J. M. Olley, IAHS Publication, 276, 443-450.
- Smith HG, Blake WH. 2014. Sediment fingerprinting in agricultural catchments: A critical re-examination of source discrimination and data corrections. *Geomorphology*, 204, 177-191. DOI: 10.1016/j.geomorph.2013.08.00.
- Smith HG, Evrard O, Blake WH, Owens PN. 2015. Preface – addressing challenges to advance sediment fingerprinting research. *Journal of Soils and Sediments*, 15, 2033-2037. DOI: 10.1007/s11368-015-1231-2.
- Stan Development Team. 2015. Stan modelling language: user's guide and reference manual. Version 2.9.0.
- Stewart HA, Massoudieh A, Gellis A. 2015. Sediment source apportionment in Laurel Hill Creek, PA, using Bayesian chemical mass balance and isotope fingerprinting. *Hydrological Processes*, 29, 2545-2560. DOI: 10.1002/hyp.10364.
- Vale SS, Fuller IC, Procter JN, Basher LR, Smith IE. 2016. Characterization and quantification of suspended sediment sources to the Manawatu River, New Zealand. *Science of the Total Environment*, 534, 171-186. DOI: 10.1016/j.scitotenv.2015.11.003.

Van den Meersche K, Soetaert K, Middelbury JJ. 2008. A Bayesian compositional estimator for microbial taxonomy based on biomarkers. *Limnology and Oceanography: Methods*, 6, 190-199.

Walling DE, Owens PN, Leeks GJL. 1999. Fingerprinting suspended sediment sources in the catchment of the River Ouse, Yorkshire, UK. *Hydrological Processes*, 13, 955-975.

Walling DE, Collins AL, McMellin GK. 2003. A reconnaissance survey of the source of interstitial fine sediment recovered from salmonid spawning gravels in England and Wales. *Hydrobiologia*, 497, 91-108. DOI: 10.1023/A:1025413721647.

Walling DE. 2013. The evolution of sediment source fingerprinting investigations in fluvial systems. *Journal of Soils and Sediments*, 13, 1658-1675. DOI: 10.1007/s11368-013-0767-2.

Ward EJ, Semmens BX, Schindler DE. 2010. Including source uncertainty and prior information in the analysis of stable isotope mixing models. *Environmental Science and Technology*, 44, 4645-4650. DOI: 10.1021/es100053v.

Accepted Article

Tables

Table 1: Summary of the four mixing model versions.

Model Version	Inference	Type	Software/ Package	MCMC Sampler	Source Uncertainty	No. MCMC Chains	Iterations	Burn-in	Jump Length
M1	Bayesian	Empirical	JAGS	Gibbs	Partial	3	750,000	250,000	225
M2	Bayesian	Empirical	Stan	No-U-Turn	Partial	3	10,000	5,000	25
M3	Bayesian	Full	Stan	No-U-Turn	Full	3	10,000	5,000	25
M4	Frequentist	Frequentist	limSolve	-	-	-	10,000	-	-

Table 2: Number of collected soil/sediment samples used to characterise the source distributions.

Set	Number of samples			
	Subsurface	Road Verge	Topsoil	Total
1	60	30	30	120
2	40	20	20	80
3	20	10	10	40
4	5	5	5	15

Table 3: Percentage source contributions for the six synthetic target mixtures

Mixture	Source Contribution (%)		
	Subsurface	Road verge	Topsoil
1	0	0	100
2	0	50	50
3	20	60	20
4	33	33	33
5	50	0	50
6	75	12.5	12.5

Table 4: Summary of mixing model performance for the six artificial mixtures.

Model Version	Mean deviation between actual and median predicted proportions (%)				Mean 95% CI width (%)			
	SUB	RV	TS	Overall	SUB	RV	TS	Overall
M1	11.1	16.0	11.5	12.9	39.4	73.6	71.0	61.4
M2	10.8	14.3	9.8	11.6	36.3	66.5	63.2	55.3
M3	9.6	14.1	10.5	11.4	41.2	66.2	61.9	56.4
M4	4.2	11.2	11.2	8.9	59.6	89.3	61.0	70.0

Table 5: Median (95% credible/confidence interval) suspended particulate matter (SPM) load contributions estimated by four mixing model versions for the entire February 2013 precipitation event.

Model Version	SPM Load Contribution (kg)		
	Subsurface	Road Verge	Topsoil
M1	386 (179–798)	562 (124–1085)	445 (34–980)
M2	428 (187–865)	357 (57–955)	604 (86–1072)
M3	427 (181–879)	351 (54–959)	608 (88–1071)
M4	326 (2–859)	407 (0–1252)	650 (183–1099)

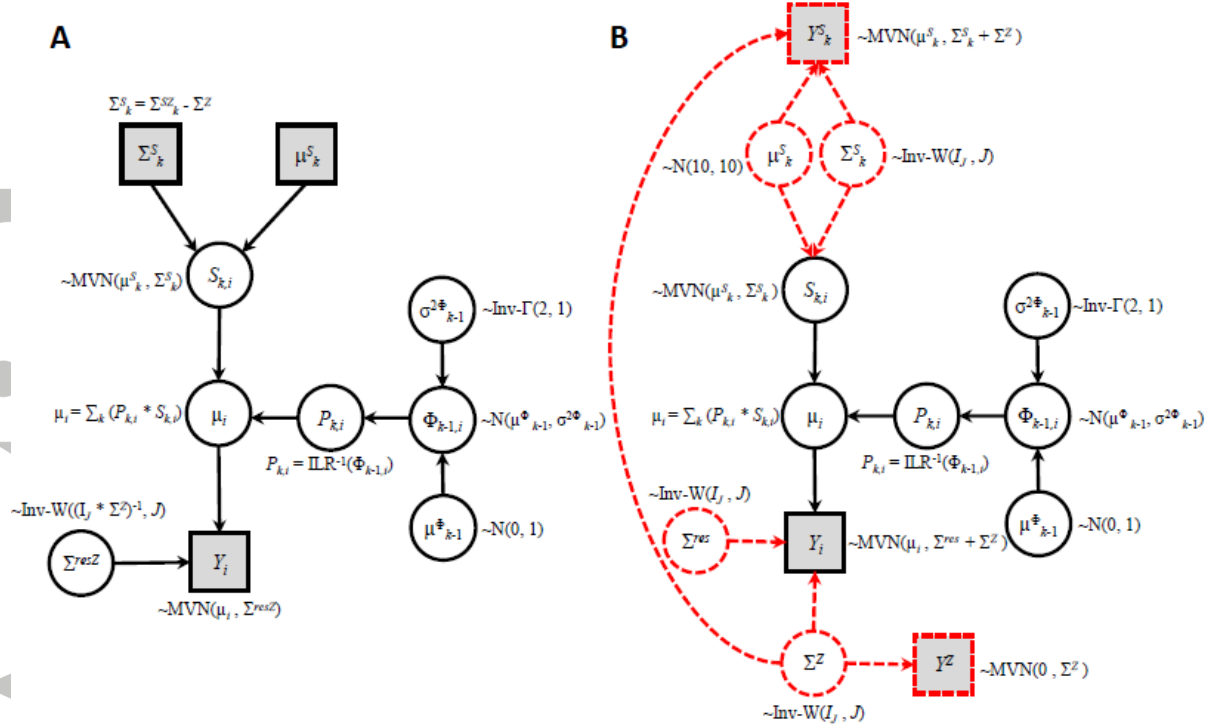


Figure 1: Directed acyclic graphs of (A) the Cooper et al. (2014a) empirical Bayes mixing model and (B) the full Bayes model introduced here. White circles denote random variables estimated by Markov Chain Monte Carlo (MCMC) sampling. Grey squares denote nodes of observed data, including parameters estimated “offline” by Maximum Likelihood in the empirical Bayes approach. Prior distributions and deterministic link equations are presented adjacent to nodes. MVN, N, Inv-W and Inv- Γ represent multivariate normal, normal, inverse Wishart and inverse gamma distributions, respectively. Dashed red lines represent the extension to full Bayes.

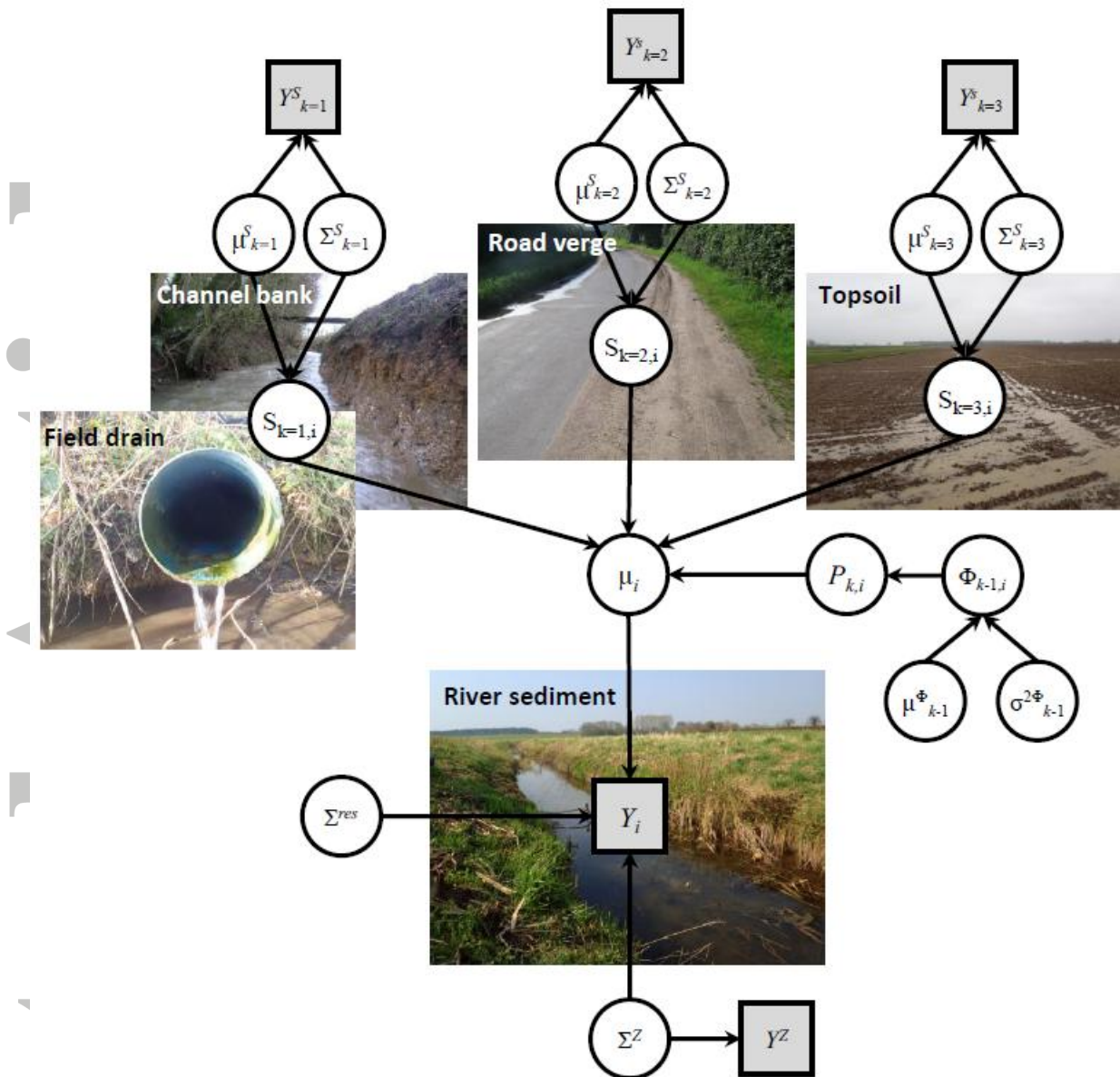


Figure 2: Schematic of the sediment source areas identified across the River Blackwater catchment, with overlay of the extended full Bayesian Directed Acyclic Graph (DAG).

Accept

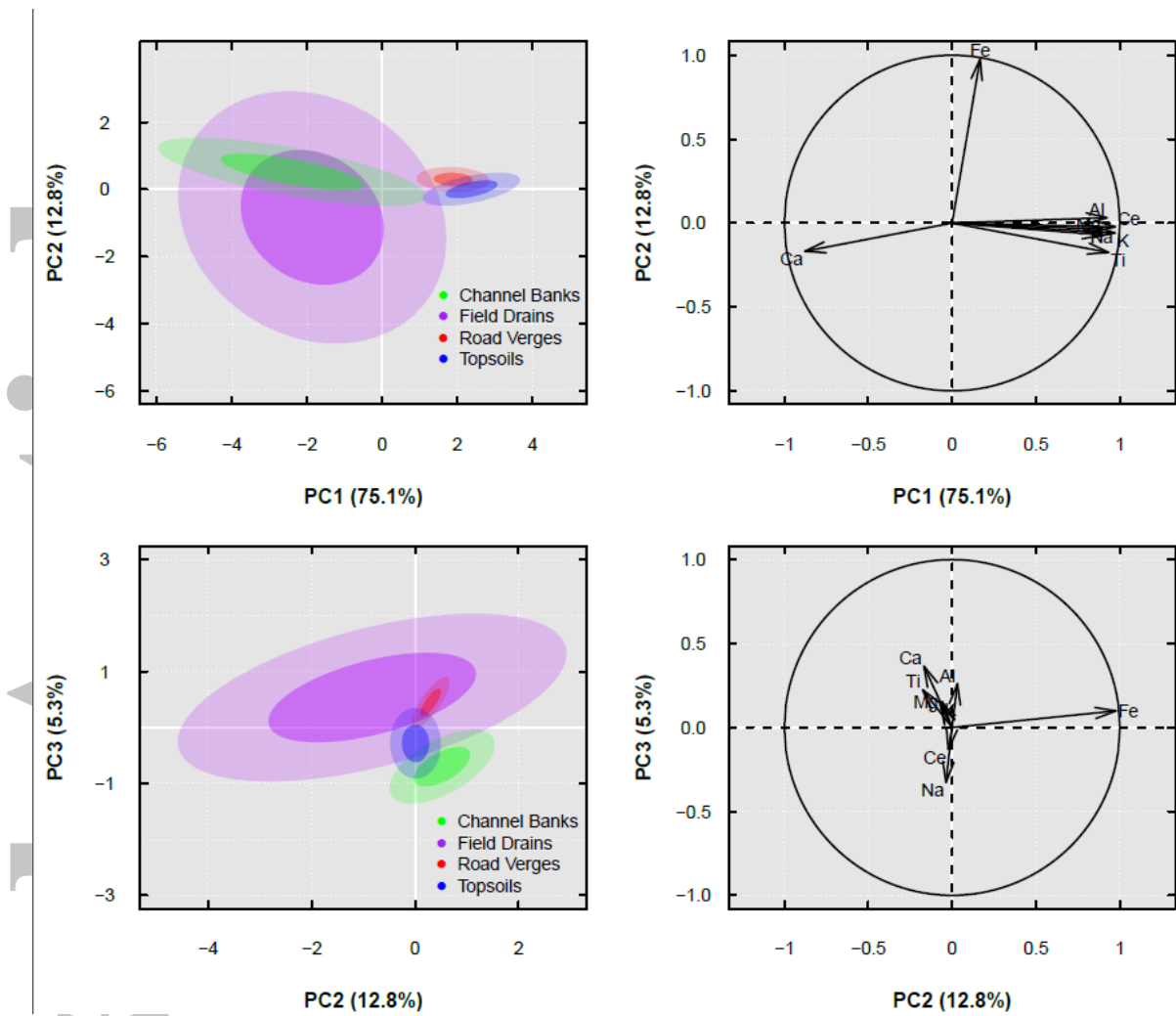


Figure 3: Principal component analysis plots of the source area samples (left) and fingerprint loadings (right) for the first three principal components. Shaded ellipsoids encompass 50% and 95% of the source area range.

Accepted

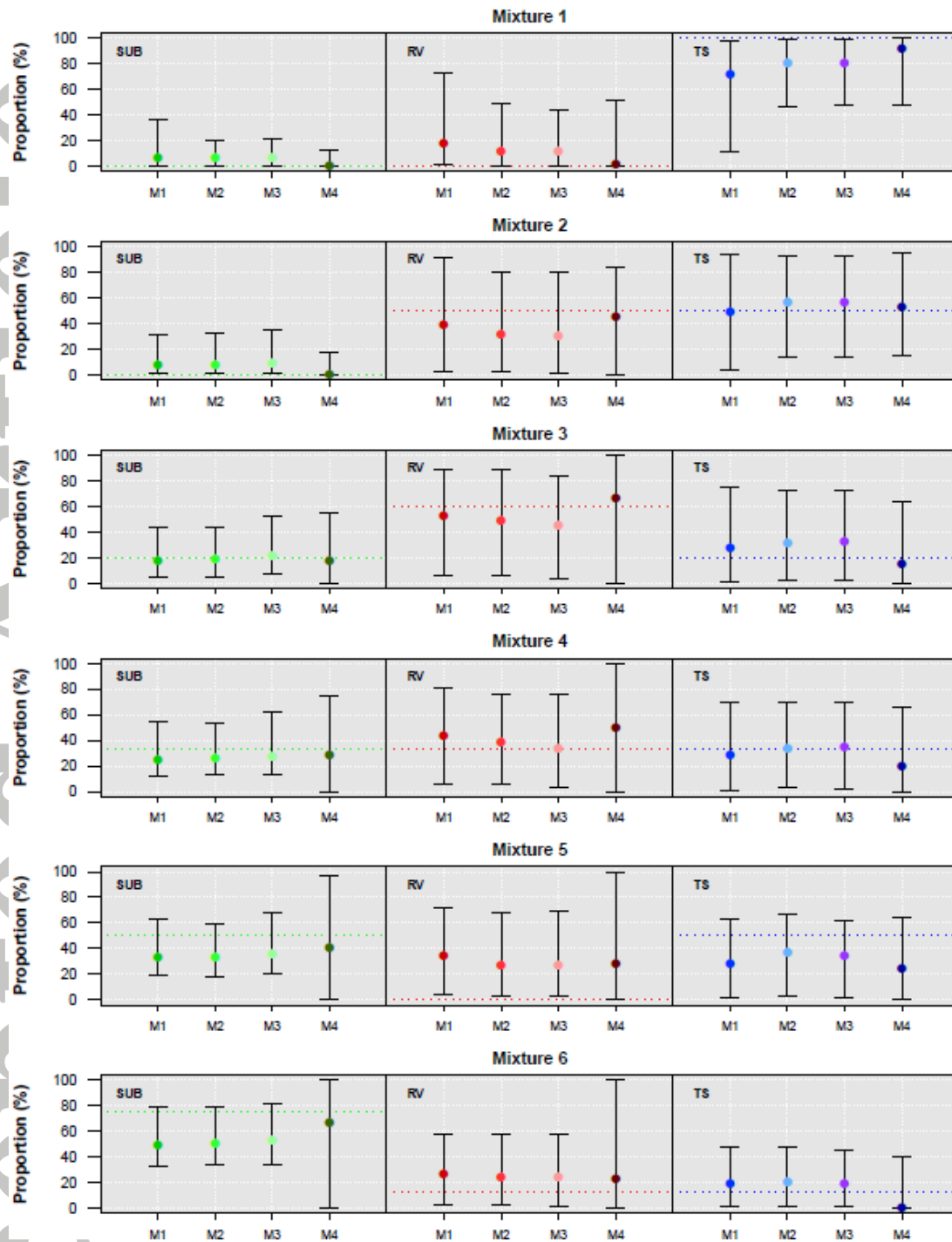


Figure 4: Comparison of sediment source apportionment results estimated by the JAGS empirical (M1), Stan empirical (M2), Stan full Bayes (M3) and frequentist (M4) mixing models for the six synthetic target mixtures. Medians (points), 95% credible/confidence intervals (error bars) and actual contributions (dotted lines) are shown.

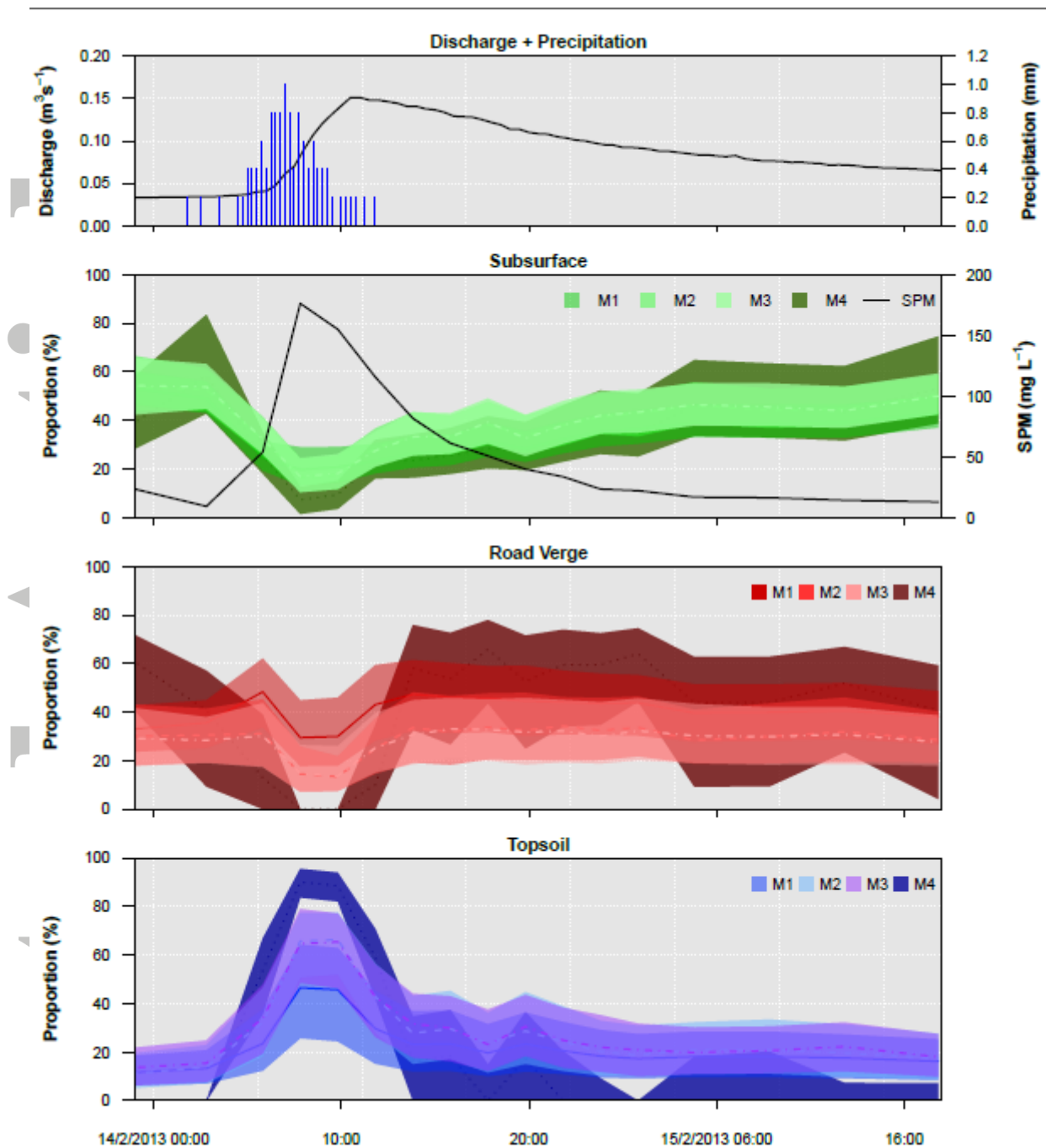


Figure 6: Effect of reducing the number of source area samples on (A) model accuracy and (B) model uncertainty. Model accuracy refers to the median contribution. Power law regression lines plotted for model accuracy; linear regressions plotted for model uncertainty.

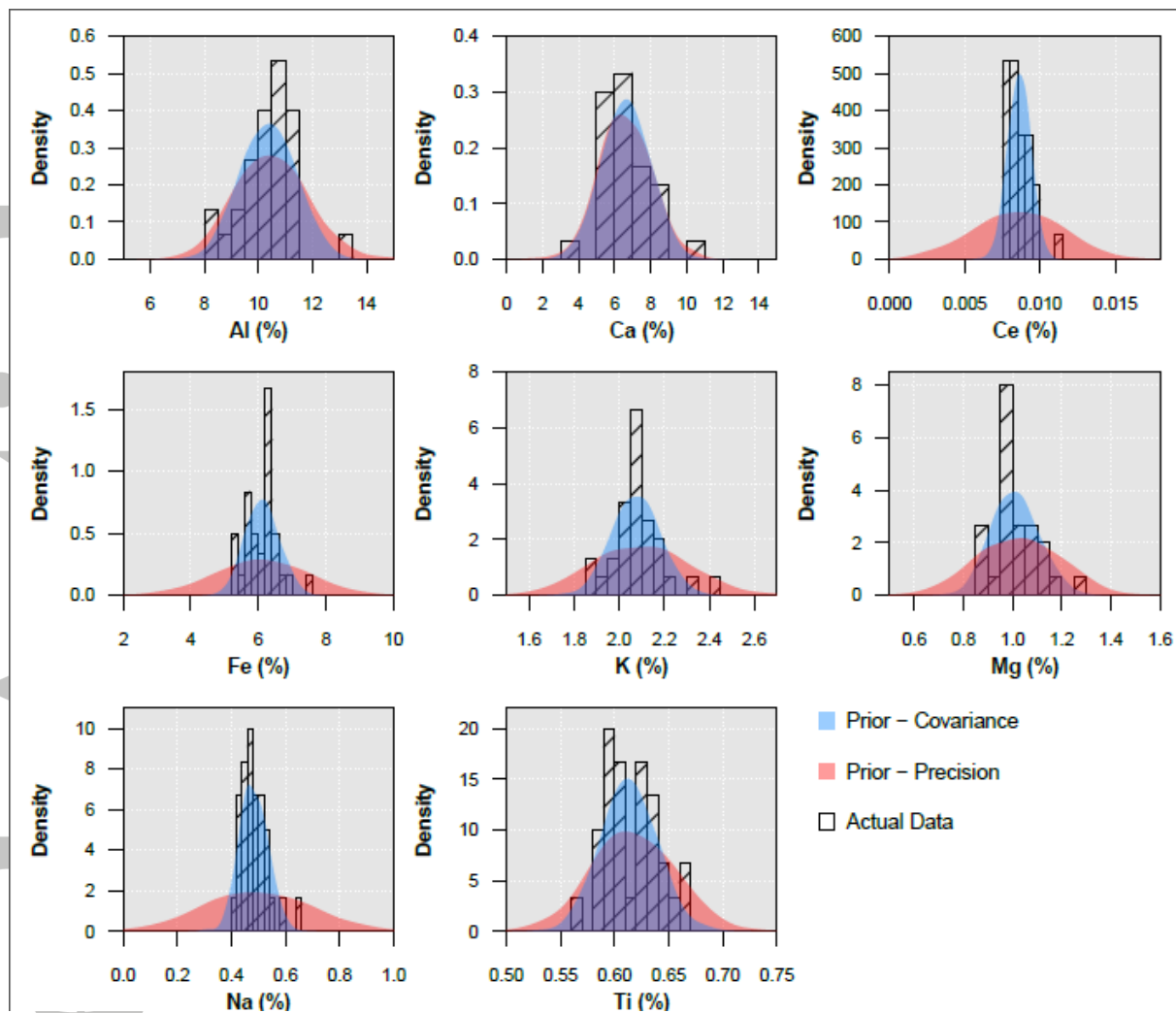


Figure 7: Histograms of the actual road verge source concentrations plotted alongside the prior multivariate normal distributions estimated using covariance (Stan) and precision (JAGS) matrices as hyper-parameters for the covariance. Prior distributions plotted using kernel density smoothing.

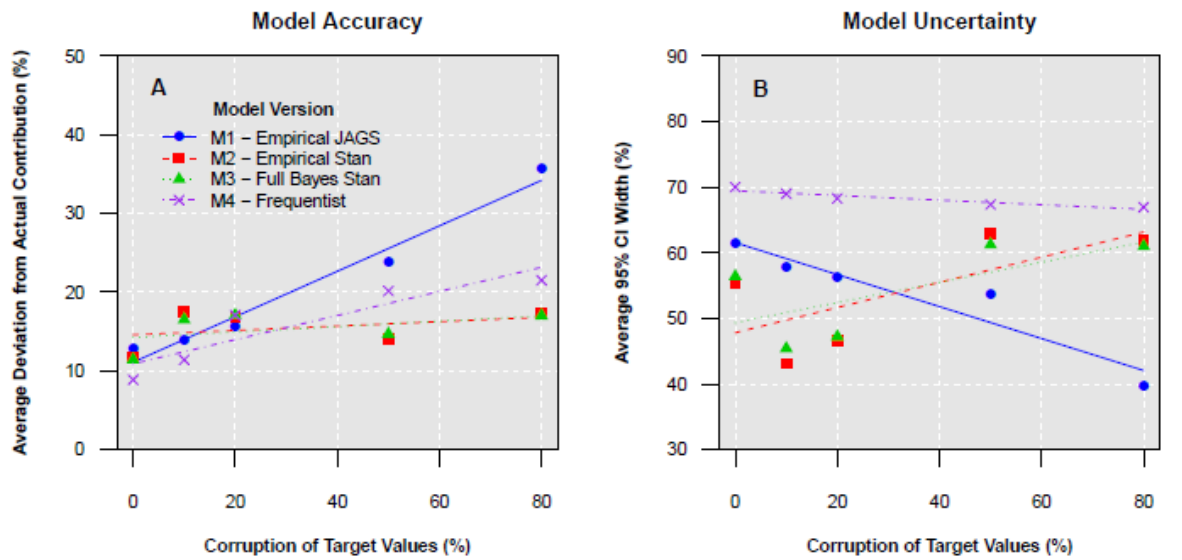


Figure 8: Impact of corrupting the target geochemistry data to simulate non-conservative behaviour on (A) model accuracy and (B) model uncertainty. Model accuracy refers to the median estimated contribution. Linear regression lines plotted for both model accuracy and uncertainty.

Accepted