Aggregated monitoring of enhanced weathering on agricultural lands

Tim Jesper Suhrhoff[1,2], Anu Khan[3], Shuang Zhang[4], Beck Woollen[1,3], Tom Reershemius[5], Mark A. Bradford[1,6], Alexander Polussa[1,6], Ella Milliken[2], Peter A. Raymond[6], Christopher T. Reinhard[7], Noah J. Planavsky[2,1]

1: Yale Center for Natural Carbon Capture, New Haven, CT, USA
2: Department of Earth & Planetary Sciences, Yale University, New Haven, CT, USA
3: Carbon Removal Standards Initiative, Washington D.C., USA
4: Department of Oceanography, Texas A&M University, College Station, TX, USA
5: School of Natural and Environmental Sciences, Newcastle University, Newcastle-upon-Tyne, UK
6: The Forest School, Yale School of the Environment, Yale University, New Haven, CT, USA
7: Department of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

Abstract: Terrestrial enhanced weathering (EW) on agricultural land is a promising carbon dioxide removal (CDR) pathway with high potential to scale. Enhanced weathering also has the potential to provide significant agronomic co-benefits to farmers and producers. Today, most EW field trials are funded through the voluntary carbon market (VCM) with the purpose of generating carbon removal credits for corporate sustainability goals. As a result, monitoring, reporting, and verification (MRV) frameworks for EW are designed for attribution of tons of removal via weathering to individual fields. Here, we describe approaches for aggregation of weathering indicators across multiple fields using aqueous, solid, and gas phase measurements. First, we demonstrate that larger agricultural catchments are at least as suitable as smaller ones for detecting weathering signals in river chemistry, and in some cases may even offer advantages due to lower variability and background weathering fluxes. Second, we assess quantification uncertainty from in-field solid phase soil measurements at increasing scales and show that errors in CDR quantification can be reduced by aggregating signals over many fields. Third, we expand our in-field analysis to consider the cost-uncertainty trade-offs of in-soil gas flux monitoring at scale. Critically, we also highlight that aggregation sets must be defined in advance and all plots included, as biased selection of fields can generate apparent removal signals out of statistical noise. Taken together, we find that aggregated monitoring of EW—quantifying CDR over multiple fields at once—can both improve existing MRV frameworks and support integration of EW practices with a broader array of government policies, unlocking funding and public support to achieve climate-relevant scale.

1 Introduction

Deep and immediate emissions reductions are needed to prevent the worst harms of climate change (UNEP 2024, IPCC 2018). In addition to emissions mitigation, there is growing scientific and political consensus that atmospheric carbon dioxide removal (CDR) will be necessary to stay within the temperature targets of the Paris Agreement (Luderer *et al* 2018, Rogelj *et al* 2018, IPCC 2022). A range of carbon removal solutions are likely needed to achieve the gigaton-scale drawdown of carbon from the atmosphere proposed in net-zero and net-negative (overshoot) climate scenarios (IPCC 2022, Geden *et al* 2024, Lamb *et al* 2024) based on local energy, land, infrastructure and mineral resource availability. One promising approach for CDR is terrestrial enhanced weathering (EW) on agricultural land. In this process, crushed cation-rich rocks applied to fields react with dissolved atmospheric carbon in water, forming aqueous bicarbonate ions. Carbon is durably stored (longer than 10,000 years; Renforth and Henderson 2017) in the ocean as bicarbonate or in soils and sediments as solid carbonate. Global removal potential via EW is estimated to be 0.5- 2 Gt/yr per year (Beerling *et al* 2020), or 64-217 gigatons cumulatively by 2080 (Baek *et al* 2023), meeting up to ~20% of expected CDR needs. Enhanced weathering can also deliver agronomic benefits, including soil pH management and improved crop yields (Levy *et al* 2024), increasing the likelihood of adoption of EW practices and projects in alignment with climate goals.

Growing demand for carbon removal services, particularly from forward-looking buyers in the voluntary carbon market (VCM), has driven a rapid increase in investment in EW companies and projects. Over 25 EW companies are operating globally and ~600,000 tons of carbon removal credits have been sold, though only a small fraction (<2%) have been delivered (CDR.fyi 2025). This trend is promising but raises important questions about the ability to accurately estimate net carbon removal from the atmosphere via EW. This is particularly true when project outcomes (e.g. carbon credits delivered) are used to make investment decisions and, ultimately, claims about net-zero targets, the primary use case of carbon removal projects to date. Historical precedent suggests that the incentive structure of the VCM drives towards low costs and inflated claims and has led to systematic failures in quantification (Gill-Wiehl *et al* 2023, Badgley *et al* 2022, Sanders-DeMott *et al* 2025) without rigorous monitoring of project impacts. Moreover, a growing body of evidence from soil organic carbon quantification (and carbon credit issuance) suggests that measurement across multiple fields (Bradford *et al* 2023, Potash *et al* 2025), using regional baselines (Oldfield *et al* 2022), can improve the accuracy of estimated carbon sequestration.

Achieving large-scale, verifiable CDR via EW will require balancing accuracy and cost, using a mix of methods (gas, solid, and liquid phase measurements, as well as models) with context-appropriate study design, across spatial and temporal scales (Clarkson *et al* 2024, Almaraz *et al* 2022). Here, we explore the benefits of monitoring EW over multiple fields or a region - rather than treating individual field estimates as accurate - through the physical aggregation of weathering products in streams and the statistical aggregation of field-level data points, to reduce uncertainty

in carbon quantification. We also discuss the cost and scalability implications of these findings from both market and policy perspectives.

2 Methods

This study evaluates the potential of aggregated monitoring—that is, quantifying carbon removal by monitoring multiple field sites together—of EW on croplands. We focus on the application of silicate materials and consider both in-field and downstream signals of weathering products, including cations and carbon species. To illustrate the utility of aggregation, we analyze monitoring approaches across three domains—streams, soils, and gas fluxes—using a series of simple models that capture first-order system behavior. These models are not intended to provide precise forecasts but rather to demonstrate how aggregation can improve detectability and reduce uncertainty in EW monitoring.

In framing this analysis, it is worth distinguishing between two related but distinct challenges: (i) reducing statistical noise from spatial heterogeneity, for which aggregation across multiple fields or watersheds provides a tractable solution, and (ii) addressing system-level processes that can alter the effective permanence or transport of weathering products, which may require moving beyond near-field soil measurements to downstream integration in rivers or groundwater. Recent work has shown for example that solute export reflects not only soil-scale weathering reactions but also subsurface redox structure, mineral buffering, and hydrological residence times (Shaughnessy and Brantley 2023, Shaughnessy *et al* 2023), emphasizing the need to capture processes that integrate across critical zone compartments. In this study, we attempt to address (i) through analysis and simulations of water, soil, and gas datasets and processes, whereas (ii) is inherently captured only in the stream water approach, since by definition this medium integrates across subsurface transport, mineral buffering, and hydrological residence times.

2.1 Watershed analysis

Enhanced weathering can be quantified by tracking weathering products such as cations or alkalinity in the aqueous phase (Clarkson *et al* 2024, Almaraz *et al* 2022, Sutherland *et al* 2024). Tracking bicarbonate alkalinity in effluent water has the benefit of quantifying CDR after losses due to formation of secondary phases and other potential means for carbon loss within soils and upstream portions of rivers. Cation fluxes are also often used as a proxy for alkalinity (Bijma *et al* 2025). However, this method of quantification introduces the challenge of detecting changes in riverine fluxes compared to baseline. Nevertheless, multiple studies have detected significant changes in riverine chemistry through time due to agricultural liming (Hamilton *et al* 2007, Oh and Raymond 2006, Barnes and Raymond 2009, Duan *et al* 2025), providing an example of the feasibility of this approach.

Here we assess the feasibility of tracking weathering products in downstream waters as a method of estimating carbon removal from the large-scale field-application of silicate minerals. First, we use a large river database from the continental US (USGS 2016) to establish baseline riverine fluxes. To ensure baseline quality, we only considered stations with at least 10 years of baseline

data since 1990, where each year has at least 10 measurements distributed over all four seasons (n=95 sites for alkalinity, n=81sites for Ca, n=80 sites for Mg, 120 unique stations). We furthermore filtered out watersheds with less than 10% agricultural area (cropland and pasture; USGS 2024), leaving n=50 sites for alkalinity, n=26 sites for Ca, n=26 sites for Mg, and n=51 unique sites in total. The locations of these stations are displayed in Figure 1a. Second, we calculate the increase in alkalinity, Ca, and Mg concentrations needed to produce a 2 standard deviations (σ) change relative to annual average baseline concentrations, assuming that a 2σ increase can be reliably detected. Third, the increase in riverine fluxes required to detect CDR is translated into an agricultural-area normalized basalt application rate based on US-average basalt composition (Lehnert *et al* 2000), watershed catchment size (USGS 2016), and areal extent of agriculture in each catchment (cropland and pasture area) (USGS 2024). For alkalinity, this is calculated based on charge balance from basalt base cation content. Note that some watersheds are nested within each other, not all datapoints are independent.

The aim of this analysis is not estimating the exact basalt application rates whose dissolution could be detected, but rather to assess how the utility of stream-based MRV scales with catchment size. This analysis is built on a steady state assumption, which implies that (1) rock added dissolves quickly, (2) weathering products are transported quickly through into streams, and (3) that weathering products are transported in river water in a way that can be detected in downstream stations. Each of these assumptions can be challenged (Kanzaki *et al* 2025, Kirchner and Neal 2013, Godsey and Kirchner 2014, Calabrese *et al* 2022, Power *et al* 2025) and the *absolute* values this analysis arrives at certainly underestimate required basalt application. However, we argue that these limitations should not scale with catchment size and that the *trend* of required basalt application, dissolution, and transport rates can nevertheless be informative to gauge the utility or river-water MRV at scale – see section 4.1 for a more detailed discussion of these limitations.

2.2 Soil analysis

An alternative way to quantify CDR via EW is the use of soil mass balance approaches (Kantola *et al* 2023, Reershemius and Suhrhoff 2023, Reershemius *et al* 2023, Clarkson *et al* 2024, Suhrhoff *et al* 2024, 2025) to assess the difference in feedstock concentration before and after weathering, by comparing changes in the total amount of mobile cations (e.g. Ca^{2+}, Mg^{2+}) in a soil sample relative to an immobile tracer (e.g. Ti.) Combined with other measurements and assumptions (Campbell *et al* 2023, Reershemius *et al* 2023, Clarkson *et al* 2024, Suhrhoff *et al* 2024, 2025), this dissolution fraction can act as a proxy for CDR. Here, we demonstrate that aggregating over multiple fields notably increases the accuracy of soil-based approaches to quantify rock powder dissolution, drawing from similar analysis for soil organic carbon accrual due to changes in land management practice (Potash *et al* 2025, Bradford *et al* 2023). We highlight that our approach here tests the ability to overcome high within- and among-field variability in cation concentrations through monitoring approaches.

We use existing data for the elemental composition of a large number of agricultural fields (Smith *et al* 2013) and basalt rock powder (Lehnert *et al* 2000). We also present recently generated data on in-field variance of elemental compositions resulting from spatial heterogeneity and analytical errors from 5 densely sampled field sites (Suhrhoff et al 2025; see supplement S2). Based on this data, we conduct a Monte Carlo simulation to assess the expected average error on detected dissolution fractions when using soil mass balance. Simulations are conducted at multiple rock powder application amounts (total amounts of 50 and 100 t ha^{-1}; not annual rates), dissolution fractions (0.25 and 0.5), and spatial sampling frequencies (1 to 20 samples ha^{-1}), where sampling frequency can be achieved either through independent point samples or by pooling sub-samples is subsamples are spaced far enough to capture in-field heterogeneity (ITRC 2020, Clausen *et al* 2013b, 2013c, 2013a). For each simulation, the accuracy of detected dissolution fractions compared to the simulated known value is compared (1) based on the average absolute error for each individual field (corresponding to single-field MRV) and (2) after averaging over multiple (10, 50, and 100) fields. Simulations are then repeated 100 times, and we calculate for both approaches the average error on quantified dissolution fractions and the frequency of overestimating rock dissolution by more than 20%. The model assumes non-paired sampling approaches where baseline and post-weathering samples are taken at random locations–though for comparison we also model paired sampling. The workflow of the simulations as well as underlying data and assumptions are explained in detail in supplement S2 (see Figure S8 for a flow chart). All code can be found in the supplement.

2.3 Gas analysis

Enhanced weathering is commonly described as capturing atmospheric CO_2 and converting it to bicarbonate. The more precise way to describe EW, however, is that it decreases the flux of CO_2 from soil that was produced largely by root respiration and degradation of organic matter. It follows that one of the most direct ways of tracking field EW CDR rates is by monitoring changes in CO_2 fluxes coming from soils (Dietzen *et al* 2018, Stubbs *et al* 2022, Rausis *et al* 2022, Yan *et al* 2023, Vienne *et al* 2024, Kantola *et al* 2023). Embedded CO_2 sensors can be used to reconstruct gas fluxes from soils using the gradient method (Maier and Schack-Kirchner 2014). Initial work on this method in a basalt trial in the Southeastern US suggests that, with continuous data acquisition, it may be an effective means of estimating relative changes in carbon fluxes (Milliken *et al* 2025).

Here we apply the same framework as solid-phase soil analysis to monitoring of CDR in-field via soil gas phase sensors. However, due to the current cost constraints (i.e., >1,000$ per sensor pair), we model much lower sensor densities and assess accuracy as a function of aggregation (number of grouped fields) to gauge if aggregation may make such an approach feasible at scale. For each simulated deployment we assume soil CO_2 emissions of 3.5 t ha^{-1} yr^{-1} for control sites, representative of soils in corn-based cropping systems (Kazula and Lauer 2023). Due to spatial and temporal heterogeneity in gas fluxes, we assume a standard deviation of 1.5 t ha^{-1} yr^{-1} on such

measurements. We simulate treatment sites by deducting nominal amounts of CDR (0.1, 0.2, and 0.5 t ha^{-1} yr^{-1}) from control site emissions. Both the chosen variance as well as CDR rates detectable based on in-soil gas sensor MRV approaches are conservative compared to observed values (Kazula and Lauer 2023, Milliken *et al* 2025) and chosen specifically to not overestimate the utility of such a low-density sensor approach. To account for the fact that EW interventions likely increase variance of emissions (due to heterogeneous feedstock application, impacts on soil organic carbon, and changes in productivity) we add the nominal CDR rates to the assumed standard deviations (such that they are 1.6, 1.7, and 2.0 t ha^{-1} yr^{-1}, respectively). From these distributions, we generate sets of control-treatment pairs, and calculate the detected CDR. We then average over increasing numbers of deployment sites (up to 10,000) and calculate the average error on detected CDR as well as the frequency with which CDR is overestimated by more than 20%. First, we simulate average CDR error and overestimation rates for 1 sensor pair per deployment (control and treatment) irrespective of field size. Second, we simulate error and overestimation for deployments with 1 sensor pair per 10 ha, assuming again field sizes between 10 and 100 ha.

3 Results

3.1 Watershed analysis

Our analysis indicates that basalt addition is more detectable (i.e., requires lower application rates) in watersheds with larger total agricultural area (Figure 2). Required basalt application rate to cause a 2σ increase compared to baseline river concentrations forms a significant trend for alkalinity ($p < 0.001$) though at low R^2 (0.25). There is no significant trend with catchment size for Ca and Mg. Ignoring loss and lag processes due to slow weathering and retardation of weathering products, the average required basalt dissolution and transport rate (in tonnes per hectare per year) for watersheds with more than 1 km^2 of agricultural area and more than 20% agricultural land cover (crop and hay/pasture) (USGS 2024) are 0.63 ± 0.68 (1σ) for alkalinity (n=45), 0.97 ± 0.87 (1σ) for Ca (n=23), and 0.53 ± 0.31 (1σ) for Mg (n=23).

3.2 Soil analysis

Aggregating soil mass balance approaches over multiple fields increases their robustness in quantifying CDR from EW. The primary measures deployed for robustness here are the average error over all model simulations, where error refers to the difference between the calculated rock powder dissolution fraction (τ_j) based on the Monte Carlo simulations and the assumed, simulated value (Figure 3). We also calculate the frequency with which rock powder dissolution is overestimated by more than 20% (Figure 3b), as an indicator of the risk of generating excess carbon credits when EW projects are developed for the voluntary carbon market. Both error and overestimation frequency are much lower when aggregating over multiple fields. Increasing the number of fields leads to a progressively smaller incremental increase in robustness. By aggregating over multiple fields, the average error at a sampling frequency of 10 samples ha^{-1} (100 t ha^{-1} basalt application, τ_j of 0.25) can be reduced from >20% for individual fields to less than 10% (10 fields) and to 5% (100 fields). The impact on the over-crediting frequency is even more extreme, where aggregating over multiple fields collapses the frequency from >20 % for individual fields to essentially zero when aggregating over at least 50 fields at the same sampling frequency.

3.3 Gas analysis

The average error on detected CDR from low-density in-soil gas sensor measurements decreases with increasing aggregation (number of fields) as well as observed CDR (Figure 4). Based on an individual sensor pair (for field sizes from 10 to 100 ha), the average error is >10 % even when 1,000 fields are grouped and at 10,000 fields <10% only for 0.2 and 0.5 t ha^{-1} yr^{-1}. The frequencies of overestimating CDR by at least 20% are approximately 10, 25, and 50% at 1,000 fields, and 0, 5, and 20% at 10,000 fields for nominal CDR rates of 0.5, 0.2, and 0.1 t ha^{-1} yr^{-1}, respectively. Using a sensor density of 1 sensor per 10 ha, both average error and CDR overestimation frequency decrease. In this case, at 10,000 fields, all nominal CDR rates are detected on average with less than 10% error. Carbon dioxide removal overestimation frequencies are approximately 0, 15, and

238 25% at 1,000 fields, and 0, 0, and 5% at 10,000 fields for nominal CDR rates of 0.5, 0.2, and 0.1 t
239 ha^{-1} yr^{-1}, respectively.

4 Discussion

4.1 Watershed

Our analysis indicates that within the existing set of USGS stream gage stations, detectable EW signals are impacted by catchment agricultural area (Figure 2) or catchment size (Figure S1) to some extent. For alkalinity, we find a statistically significant but relatively weak relationship between required basalt application rates and total agricultural area ($p < 0.001$, $R^2 = 0.25$; Figure 2), indicating that less basalt may be required in larger watersheds for viable stream water MRV. However, there are no clear trends for Ca and M, suggesting that signal detectability is shaped by a combination of factors.

One factor likely contributing to the significant trend for alkalinity appears to be that weathering rates (i.e., area normalized alkalinity fluxes) tend to be lower in large agricultural watersheds ($p < 0.05$ but low R^2 of 0.09; Figure 5a; see Figure S5 for total-catchment area normalizations and Figure S6 for Ca and Mg data). This seems at least partially related to lower runoff with increasing catchment area ($p < 0.001$, $R^2 = 0.35$; Figure 5b), runoff being a widely recognized control on weathering rates (White and Blum 1995, Gaillardet *et al* 1999, Gislason *et al* 2009, Hartmann 2009, West 2012). By contrast, we do not observe a significant relationship with erosion rates ($R^2 = 0.03$, $p = 0.19$; Figure 5c). One additional factor that may help explain the relationship in Figure 2a–i.e., lower required basalt application in larger agricultural catchments–could be that variability in stream chemistry appears to decline with catchment size, as suggested by the significant decrease in the relative standard deviation of annual baseline data ($p < 0.001$, $R^2 = 0.21$; Figure 5d).

Taken together, these observations suggest that larger agricultural watersheds are at least as suitable—and in some cases more advantageous—than small ones for aggregated monitoring of EW. When new monitoring stations are established directly within agricultural catchments, watersheds of different sizes may be similarly well suited for signal detection (see Figure S7 for stations with >50% ag-land fraction). While this conclusion is necessarily constrained to the specific set of USGS stations analyzed here, it highlights that large catchments can be promising candidates for steam-based MRV. The added advantage of large catchments is that the same infrastructure can monitor broader agricultural areas, lowering MRV costs per ton (see Section 4.4), and making them particularly attractive for future deployment.

We use this signal-to-noise analysis primarily to explore how area-normalized basalt application rates may relate to catchment size, with the aim of comparing the relative utility of different watershed contexts for monitoring. We do not aim to provide definitive estimates of the absolute application rates required for signal detection. The analysis presented here is based on a steady-state assumption: the basalt that is added to fields dissolves immediately, and the dissolution products are transported through soils into streams where they impact river chemistry. We have

here neither modelled weathering processes as well as loss processes in soils such as due to the formation of secondary phases, uptake onto cation exchange sites, inclusion into biomass, or strong acid weathering (Clarkson *et al* 2024), nor lag times between weathering and the arrival of dissolution products because of interactions with exchangeable acidity (Kanzaki *et al* 2025). Hence, the analysis here should not be interpreted in terms of the absolute values indicated to be feasible from a signal-to-noise perspective, with realistic required rates certainly exceeding the 0.5-1 t ha^{-1} yr^{-1} found here. What we primarily want to demonstrate here is that—provided river chemistry is going to be useful to monitor EW given all loss and lag processes—likely it is going to be as feasible to accurately detect the same area-normalized application rates in catchments with large areas (and proportions) of agricultural land. That said, the feasibility of detecting weathering signals at watershed scale is supported by historical observations, such as the documented increase in alkalinity in the Mississippi River following widespread liming in its basin, or responses to river chemistry to wollastonite application in the Hubbard Brook catchment (Hamilton *et al* 2007, Oh and Raymond 2006, Barnes and Raymond 2009, Duan *et al* 2025, Shao *et al* 2016, Taylor *et al* 2021).

One limitation of our analysis is that it assumes that weathering products generated in soils are efficiently transported into streams, irrespective of catchment size. In reality, larger watersheds may be more prone to transport limitations, where solute export is constrained by hydrological residence times or saturation of subsurface pathways. In such cases, even if weathering fluxes increase with basalt application, the resulting products may be retained within soils, groundwater, or riparian zones rather than exported to surface waters, potentially dampening the signal relative to our trend analysis. This possibility could be reflected in the increasing baseline concentrations observed in some larger catchments ($p < 0.05$ but low R^2 of 0.11; Figure 5e), which may indicate slower turnover of solute reservoirs. These scale-dependent effects are not captured in our steady-state framework, but their impact can be mitigated by careful deployment choices, as laid out below.

In addition to possible loss processes, lag of weathering products through soils and subsurface pathways can also complicate interpretation of absolute application rates required for signal detection. However, resulting lag times are not expected to systematically scale with catchment size, such that the general trends should nevertheless be valid provided export limitation does not significantly increase with watershed scale. Within the dataset considered, watersheds with greater agricultural area do not have a longer average distance to the closest river (Figure 5f), supporting the assumption that the likely overall transport rate limiting step—cation transport through topsoils (Kanzaki *et al* 2025) and from topsoils to rivers—should not increase with catchment size. If stream water is going to be a useful medium to quantify CDR from EW, within the existing network of USGS gauges, based on the signal-to-noise analysis of alkalinity fluxes, larger streams are hence more likely to yield a clear signal at equal area normalized application rates.

The adverse impact of these loss processes and related lag times on riverine signals can be minimized by deployment choices. A number of studies based on short-term (generally <6 months) core experiments have demonstrated the impact of cation storage in secondary phases on effluent water composition (Renforth *et al* 2015, Pogge von Strandmann *et al* 2022, Iff *et al* 2024, te Pas *et al* 2025, Vienne *et al* 2025). Model results suggest that lag times due to interactions with the soil exchangeable cation pool can for example exceed 30 years in the US corn belt (Kanzaki *et al* 2025). However, this process is primarily controlled by the cation exchange capacity (CEC) as well as base saturation of soils, such that overall, the tropical soils of the south-eastern US are generally expected to have lower associated lag times (Kanzaki *et al* 2025). Similarly, precipitation of secondary carbonates is favored by high soil pH and carbonate content, both of which are generally higher in the western half of the US (Smith *et al* 2013, Wieczorek 2019), such that related losses are likely lower in the east, demonstrating the potential to reduce expected lags and losses by choosing suitable catchments (e.g. low CEC, high base saturation, high water infiltration; Kanzaki et al 2025). Loss processes are less likely to occur in catchments with high runoff and short hydrological residence times, although in practice deployment decisions will need to balance trade-offs among these parameters to identify the most suitable locations. In the context of watershed monitoring, it should be generally true that watersheds favorable for EW in terms of low loss processes and lag times are also the most suitable for detecting weathering signals downstream, demonstrating the potential to reduce expected lags and losses by choosing suitable catchments.

In summary, our analysis does not provide evidence that larger watersheds are less effective than smaller ones for stream-based MRV at equal application rates. On the contrary, they appear often to be more suitable, with the added benefit that monitoring infrastructure can cover broader agricultural areas, lowering MRV costs per ton. Provided that EW signals can be reliably detected in streams—a conclusion supported by previous work showing measurable riverine responses to liming and mineral additions (Hamilton *et al* 2007, Oh and Raymond 2006, Barnes and Raymond 2009, Duan *et al* 2025, Shao *et al* 2016, Taylor *et al* 2021)—this approach offers strong potential for monitoring EW at scale.

4.2 Soil

In statistics it is well known that individual-level “noise” typically obscures intervention effects, whereas representative sampling of individuals from a population can provide a robust and accurate average treatment effect (Holland 1986, Rothman *et al* 2008, Rubin 1974). We refer here to this fact as ‘aggregation’ and suggest it is a useful tool to estimate treatment effect size from a sample of a population that should be applicable to other parameters of interest for EW in agricultural settings, such as pH or base saturation. It has for example been extensively demonstrated in the context of soil organic carbon management (Potash *et al* 2025, Bradford *et al* 2023), where it has been postulated that detection may not be reliable at a field level, but aggregation may facilitate identifying population-level trends.

We demonstrate here that this general principle also applies to soil-based MRV for EW: Aggregating soil based MRV over multiple fields decreases the average error on detected dissolution fractions as well as the frequency of overestimating weathering by more than 20% (Figure 3). Crucially, within the level of variability in soil elemental composition modeled here, high levels of accuracy on CDR numbers are only achievable by aggregating over multiple fields. It is a general feature of soil mass balance approaches that detectability increases over time as cumulative applications increase and more time has passed for weathering (Suhrhoff *et al* 2024). Even when increasing feedstock application (Figure 6a&b) and dissolution fraction (Figure 6c&d), aggregation over multiple fields reduces error and reduces over crediting.

In our analysis we focus primarily on unpaired sampling, as a conservative estimate of potential errors in heterogeneous soils (Rogers and Maher 2025). For comparison, we also evaluate paired sampling (supplementary section S2.3). These results show that when paired sampling can be implemented reliably—whether through high-precision GPS or by physically marking sample locations—soil mass balance approaches may become feasible also at the individual field scale (Figure S12 ; Suhrhoff et al 2025). Importantly, one of the key advantages of aggregated monitoring is that it can achieve comparable accuracy at larger scales without depending on the success of paired sampling.

Statistical models like those explored here can be used to inform required sampling protocols given prescribed levels of average error and uncertainty. For example, if baseline data of soil heterogeneity was available for a given field or sets of fields, our approach to statistical modeling could be used to estimate expected average errors as a function of sampling frequency and field aggregation. This approach is similar to a recent preprint aiming to provide constructive guidance on sampling protocols (Rogers and Maher 2025), but in our case agnostic to what is an "acceptable" level of uncertainty. In this context, it is important that the minimum requirements of sampling protocols are assessed prior to deployment, and that all uncertainties resulting from CDR quantifications are properly propagated, for example via Monte Carlo simulations (Derry *et al* 2025). While we have only modeled "sampling frequency", provided that field-scale heterogeneity is captured (i.e., sub sample radius > wavelength of in-field heterogeneity), the required sampling frequency can also be achieved by pooling sub-samples. In this context, a large body of literature exists on how to accurately sample heterogenous media from the context of soil pollution remediation (i.e., incremental sampling methodologies; (Clausen *et al* 2013b, ITRC 2020, Clausen *et al* 2013c, 2013a, Hadley *et al* 2011, Hewitt *et al* 2007).

While our modeled results indicate that clear gains for MRV robustness may be achieved by aggregating over relatively small numbers of fields (e.g., 10), in practice, generating an accurate aggregate value for CDR presents several challenges. Given the unreliability of individual field-level estimates, translating into CDR is not as simple as multiplying, for each field, τ_j with application amounts per area, the field size, and assumed carbon losses (incl. LCA emissions). When using arbitrary sets of fields, such a simple weighting approach risks making the aggregated CDR result disproportionately determined by the largest fields with the most deployed rock

powder—undermining the premise of aggregation. A more rigorous approach would be to define subsets of fields that are similar in relevant respects (e.g., size, application amounts, feedstock properties, soil type, initial soil conditions such as pH and base saturation, and any parameter used for MRV and its resolvability; Suhrhoff et al 2024), so that averaging or calculating the sum of individual CDR signals within these subsets produces a meaningful aggregate signal. Defining such subsets will require much larger datasets than the pure number of fields within any one aggregated set, enabling clustering analyses and other statistical approaches that can group fields in a defensible way. Furthermore, meaningful aggregation must be achieved by grouping multiple similar-sized fields into sets, rather than artificially subdividing individual fields into smaller sub-fields, to benefit from the increase in sampling numbers that true multi-field aggregation entails. Finally, in EW projects there are typically control fields or plots, introducing additional variability and requiring still larger aggregation sets to achieve similar robustness and accuracy (Bradford *et al* 2023). In practice, unless extraordinary effort is devoted to scouting fields of comparable starting conditions, the need for large sets of fields to define meaningful aggregation subsets may naturally align more closely with frameworks suited to monitor the impact of pay-for-practice policies than with today's VCM protocols.

4.3 Gas

As with most aspects of open system monitoring, even though this is a simple concept, monitoring implementation can be difficult. Agricultural soil CO_2 fluxes—which are directly tied to management strategy, crop activity, soil moisture, and temperature—are highly variable (Barron-Gafford *et al* 2011). Change in CO_2 fluxes from soils can be detected in multiple ways. The eddy co-variance method has been used in a high application basalt trial in Illinois, the core of the corn belt in the central US (Kantola *et al* 2023). Soil chambers have also been employed in several studies, mostly in mesocosms (Clarkson *et al* 2024). These studies have had mixed results detecting signals of weathering. In-soil sensors have successfully been used to constrain CDR from EW in at least one field study (Milliken *et al* 2025). Nevertheless, it should be noted that while soil gas fluxes provide a direct link to surface–atmosphere exchange, their strong temporal variability and sensitivity to precipitation events make it challenging to integrate fluxes to annual or multi-year timescales (Hodges *et al* 2021).

Estimating in-field CDR rates using gas phase sensors typically requires a control plot with soil and environmental conditions closely matching those of the deployment area, making this method most accurate for small portions of a field, such as experimental plots within a larger deployment. The high cost of current in-soil CO_2 sensors limits sensor deployment to small areas, which undermines accuracy for project-level accounting due to spatial variability in soil CO_2 fluxes and feedstock weathering. As a result, while this approach may provide localized insights, it is unlikely to yield reliable estimates for entire fields. However, regional or jurisdictional-scale monitoring may offer more accurate field-level CDR estimates, similar to the principles applied in soil data analysis, where aggregation across a large sample set improves the accuracy of mean estimates

(Bradford *et al* 2023). Realizing this potential will require continued innovation to lower sensor costs and improved durability, enabling deployment at higher densities and larger scales.

While low-density in-soil gas sensors may not be an effective approach to monitor CDR at a field level, the analysis developed here suggests that when aggregating over many thousands of fields, it may be possible to accurately quantify CDR. The next step for evaluating whether soil CO_2 fluxes (via *in situ* sensors or gas flux chambers) can become an important component of EW MRV is estimating the scale of variability in fluxes in fields from multiple agronomic regions. This work is essential to generate an estimate of the efficacy and costs needed to accurately detect on field CDR rates using CO_2 sensors before such an approach may be applied at scale based on a sparse data foundation.

4.4 Accuracy, Cost, and Scale

Deploying EW as a climate solution requires optimization of accuracy, cost, and scale. Accuracy is necessary to ensure true climate impact. Cost reductions are needed to enable projects at prices that are palatable to corporate buyers in the voluntary carbon market and, eventually, feasible for governments to undertake as a public good at climate-relevant scale. Given that current credit prices, generated from first-of-a-kind field trials, hover around $300-400/metric ton (CDR.fyi 2025) and MRV is one of the largest cost drivers reported by CDR suppliers (Mercer *et al* 2024), novel measurement approaches are likely needed to achieve acceptable accuracy at reduced cost over time.

Watershed monitoring is one appealing option because it provides a direct, lower-bound measurement of removal rates over large areas (e.g. entire catchments) and integrates over a chain of soil chemical processes that may constitute losses of carbon from EW which are hard to measure in situ. Although individual monitoring stations can be expensive (> 20-100k plus recurring lab costs depending on instrumentation, availability of local infrastructure, and target parameters; Harmel et al 2023) including time and labor for taking samples, gage stations can cover large agriculture areas, resulting in markedly lower monitoring costs overall. Crucially, the cost of such stations would not depend on the size of the catchment, translating into lower per-area and per-ton MRV cost for larger catchments, in addition to more favorable signal-to-noise (cf., section 4.1). Gage stations can also be used for multiple socially beneficial purposes (e.g. water quality monitoring), distributing costs over a range of stakeholders.

In-field soil sampling and estimation of dissolution rates through soil mass balance is already a common practice in commercial enhanced weathering projects (Puro.Earth 2024, Sutherland *et al* 2024). Accuracy is a challenge for this method, given the high spatial variability of soil composition. Building off a similar analysis for soil organic carbon (Potash *et al* 2025, Bradford *et al* 2023), we demonstrate that averaging over 10-100 fields significantly reduces estimation error. In the context of carbon markets, this translates to reduced risk of over crediting, or generating carbon removal credits that do not reflect a real change in atmospheric CO_2

concentration. It is important to note that our results do not support trading off in-field sampling with multi-field averaging. Low in-field sampling densities (less than 2 samples/ha) consistently result in high error rates even when averaged over an increasing number of fields (Figure 3a).

An alternative approach to scaling soil-based MRV is to measure intensively on a small plot (e.g., 0.1 ha) and extrapolate those results linearly to larger deployments. However, this approach is unlikely to yield accurate estimates of carbon removal as soil composition, hydrology, and management practices vary substantially across fields, even within the same farm or watershed. This heterogeneity directly affects both the rate of weathering reactions and the detectability of their products. Empirical evidence from soil organic carbon projects shows that extrapolation from 1 to 2 small plots per field produces unreliable data because such limited sampling is unlikely to be representative of a field (Heikkinen *et al* 2013, Maillard *et al* 2017, Poeplau and Don 2015). In the context of EW, this means that a small set of intensively monitored fields could overestimate removals if applied uniformly across thousands of ha. Robust estimates therefore require approaches that capture variability across landscapes—either through direct aggregation of field measurements or integration with downstream monitoring.

It should be noted that aggregation of data from commercial deployments requires fields to be added to aggregation sets before post-treatment data is available (e.g. in the project design document), and all designated plots to be included in the final analysis for credit delivery. Due to statistical noise from soil heterogeneity and measurement error, detection of apparent CDR is expected in some cases, even when the true value is zero or the intervention is a net CO_2 source. If such low or negative detected values are selectively omitted as "noise" or the corresponding projects never delivered/reported on, the resulting average will be biased upward, leading to systematic overestimation of carbon removal (Figure 7). To avoid this, crediting of EW must move beyond individual field scales and should always be done on a set of similar deployments. Strict criteria must be set for adding or excluding plots from project or control fields during the project period, such as unexpected and unavoidable land manager change in practice. Such criteria for aggregation exist in other land use carbon crediting frameworks (e.g. criteria for exclusion of control plots from dynamic baselines in forestry protocols (Shoch *et al* 2024). Guidance for EW projects will be needed as the industry moves towards large deployments.

Lastly, we note that integration of modeling into monitoring frameworks can further optimize cost efficiency, though at present models are not a substitute for empirical approaches, neither at watershed nor field scale (Zhang *et al* 2025, Kanzaki *et al* 2025). Modeling, when paired with distributed sensor networks and targeted sampling, can strengthen robustness without dominating budgets. For example, the New York City watershed program invests ~$6.7 million annually in monitoring, of which ~10–15% supports modeling tools that provide real-time forecasting and operational guidance (NASEM 2020). For enhanced weathering, aggregated monitoring approaches that combine in-field sampling and watershed instrumentation with calibrated modeling frameworks to constrain downstream losses may provide a viable pathway for MRV at

scale and enable credible, lower-cost quantification of removals across large heterogeneous landscapes.

4.5 Policy implications

Government policy can support accurate estimation of carbon removal via enhanced weathering, at the scales discussed in this work (catchments and hundreds of agricultural fields) and beyond (eco-regions and thousands of fields within a jurisdiction and particular regulatory framework).

For watershed monitoring in the US, this analysis builds directly on publicly funded, publicly available data collected through the US Geological Survey (USGS) stream gage network. This data serves a range of purposes (both economic and environmental) and serves a diverse set of stakeholders (governments, local community-based organizations, as well as commercial entities). Though the US has a uniquely robust stream gaging network, other countries support similar public water monitoring (Barker *et al* 2022).

Expansion of water monitoring networks serves two important needs for enhanced weathering. First, before widespread deployment of EW, water monitoring enables accurate assessment of background weathering rates. This in turn enables optimal site selection for deployment and watershed pairing to assess counterfactuals. Second, after deployment, watershed-scale monitoring is the most direct measurement of carbon removal and storage in the aqueous bicarbonate reservoir, which can be used for direct estimation of removal rates as well as providing a conservative check on removal rates determined through in-field soil-based measurements.

However, if multiple projects introduce bicarbonate ions (or alkalinity) to the same waterway in the same period, fair attribution of carbon removal cannot rely solely on deployed rock amounts or treated area. This approach overlooks differences in deployment strategies that may impact net carbon storage capacity and/or losses to outgassing. A more rigorous approach would be to standardize the use of publicly available reactive transport models—expanded and cross-calibrated from existing frameworks such as SCEPTER (Kanzaki *et al* 2022, 2025, 2024), CrunchFlow (Steefel and Molins 2009), or PFLOTRAN (Mills *et al* 2007, Hammond *et al* 2007) amongst others (Taylor *et al* 2017)—to allocate watershed-scale CDR based on modeled realized fluxes. Where additional alkalinity inputs occur (e.g., from wastewater treatment or other engineered CDR methods), attribution frameworks will also need to adjust EW-derived fluxes accordingly to avoid overestimation. Watershed level monitoring of EW may require novel governance mechanisms, beyond the scope of the current VCM, to address these challenges (Woollen and Planavsky 2024).

Policy can also play a key role in supporting soil-based MRV by enabling widespread sampling and aggregation of soil data across large areas. Large-scale, publicly maintained datasets—such as those generated through national soil censuses—could provide critical baselines for site selection, monitoring, and attribution of climate impacts (Smith *et al* 2013, USGS 2024, 2023). We stress, however, that large datasets are not in themselves a panacea because robust and accurate estimation

at population scales requires representative sampling of individuals (Bradley *et al* 2021). In particular, systematic collection of soil pH and related parameters would greatly improve the accuracy of EW deployment assessments. Beyond public programs, substantial amounts of valuable soil data already exist within commercial laboratories–Waypoint (US) and the Tentamus Group (global) each analyze >1.5 million soil samples annually. Policy could unlock this resource by providing incentives for data sharing, similar to strategies used in medicine, energy, and other domains for public-private data partnerships (Susha *et al* 2023). In the US agricultural extension officers could further amplify these efforts by advising farmers on when and where to sample, ensuring that monitoring programs capture regional variability while minimizing redundancy. Together, these measures would lower barriers to building robust, aggregated soil datasets, reduce uncertainty in CDR estimates, and create shared public goods that benefit both carbon markets and broader agricultural management.

Aggregated monitoring can, in turn, enable policies that stimulate deployment of EW techniques across geographies and agricultural contexts by enabling lower cost, large (and possibly jurisdiction) scale assessment of removal fluxes. Removal fluxes can be reported in national greenhouse gas inventories towards meeting countries' Nationally Determined Contributions for the Paris Agreement. One such policy mechanism, already used across jurisdictions (CRSI 2025) are subsidies for the application of agricultural lime for soil pH management. Such subsidies, if adapted for a range of weathering feedstocks, could be structured as pay-for-practice (an area-based payment to farmers for spreading materials that remove carbon via weathering, decoupled from measurement) or pay-for-results (a base payment for spreading materials and an additional payment after verification of carbon removal via soil- and/or aqueous-phase measurement at scale).

It is important to note that the efficacy of policy to optimize soil pH management to bring about CDR will vary on a regional level and may not align with existing monitoring infrastructure. For example, there is a mismatch between the US regions (primarily in the Southeast) with highest weathering potential (Moosdorf *et al* 2011, Kanzaki *et al* 2025) and existing USGS sites with a high amount of baseline data (Figure 1a), indicating that an expansion of the current USGS river gauging network may be necessary for watershed monitoring. Moreover, use of alternative policy mechanisms (beyond carbon crediting) would benefit from consistent IPCC guidance on the accounting of EW practices in national inventories, a process that is currently underway (IPCC 2024).

In summary, there is a synergistic relationship between policy to support aggregated monitoring (e.g. expanded water quality monitoring infrastructure, public-private data sharing partnerships) and policy to support deployment of enhanced weathering. Careful consideration of CDR potential, measurement accuracy, and cost is needed to optimize across this opportunity space.

5 Conclusion

Enhanced weathering has emerged as a promising strategy for atmospheric carbon dioxide removal, yet its scalability is constrained by the accuracy, cost, and robustness of MRV. Our analysis highlights how aggregated monitoring—both through the physical integration of weathering products in streams and rivers and the statistical aggregation (i.e. averaging) of field-level measurements—can provide a pathway toward credible, lower-cost MRV at scale. By leveraging existing infrastructure and methodological advances, aggregation addresses some of the central barriers to EW deployment while aligning with practices already common in soil carbon and forestry protocols.

At the watershed level, we find no evidence that larger agricultural catchments have a lower utility of detecting EW signals in stream waters based on equal rock application, dissolution, and transport rates. In contrast, in some cases signal detection may be favorable in larger catchments due to lower variability in stream chemistry and lower background weathering fluxes. Because the fixed cost of installing and operating monitoring infrastructure is independent of catchment size, per-hectare and per-ton MRV costs are minimized in larger watersheds. This makes watershed monitoring a particularly appealing approach for large scale—possible jurisdictional—deployment, where results can serve both as direct measurements of CDR and as "top-down" validation of field-based approaches. At the field scale, soil sampling and mass-balance analysis remain among the most widely practiced MRV approaches (Clarkson *et al* 2024) in current EW projects, but our results show that accuracy is strongly constrained by spatial heterogeneity. Aggregation across multiple fields substantially increases the robustness and accuracy of soil-based CDR, reducing both the average error and the risk of over-crediting, particularly when combined with adequate in-field sampling densities. While direct soil CO_2 flux measurements using in-situ sensors are currently limited by cost and variability, the same statistical principle applies. Aggregation across larger numbers of fields or regions can transform noisy local measurements into reliable population-level estimates; though the required number of gas sensor pairs (i.e., 10k fields) may remain prohibitive of low-density gas sensor MRV approaches at acceptable accuracy. Continued innovation in low-cost sensor technology, coupled with appropriate sampling design and statistical modeling, has the potential to reduce MRV costs while improving accuracy through enabling higher density sensor networks.

Together, these findings suggest that aggregation is not simply a technical workaround but a fundamental principle for robust EW monitoring. Whether through hydrological integration at the catchment scale or statistical integration across sets of fields, aggregation lowers variance, reduces the risk of systematic bias, and enables credible estimates of CDR. Equally important, aggregation sets must be defined *a priori* and all predesignated plots—controls and treatments—must be included in analysis and crediting. Otherwise, biased selection (e.g., omitting low values or detected emissions) can manufacture apparent signals out of statistical noise and systematically overstate removals. Finally, for some MRV approaches, aggregation also shifts the cost structure

of MRV: rather than scaling linearly with the number of participating fields, costs can be amortized across larger areas and multiple stakeholders, making EW more feasible for inclusion in both VCMs and national greenhouse gas inventories.

Looking forward, further work is needed to refine the economic and policy frameworks that could support aggregated monitoring. Quantitative comparisons of MRV costs across monitoring approaches—soil, aqueous, and gas flux—will help identify optimal deployment strategies under different agronomic and hydrological contexts. At the policy level, expansion of public water and soil monitoring infrastructure, combined with protocols that mandate inclusion of all fields within predefined aggregation sets, can provide the transparency and rigor necessary for credit issuance at scale while also providing societal co-benefits. Integration with existing agricultural support mechanisms, such as liming subsidies or soil census programs, offers a ready avenue for embedding EW within established governance structures.

In sum, while uncertainty remains regarding the absolute rates of CDR achievable through EW, our results demonstrate that aggregated monitoring provides a viable path toward accurate, scalable, and cost-effective MRV. By combining physical integration in watersheds with statistical integration across landscapes, this approach can underpin the credibility of EW as a climate solution that brings real benefits to farmers, while simultaneously lowering costs and facilitating policy uptake. As deployment scales up, aggregation may prove to be the key enabling principle that bridges the gap between scientific rigor, economic feasibility, and climate impact.

6 Acknowledgements

The authors would like to thank Prof. Dr. Susan L. Brantley, Dr. Abby Lunstrum, and Dr. Lucia D. Simonelli for helpful feedback and discussion of this manuscript.

7 Funding

TJS acknowledges funding by the Swiss National Science Foundation (Grant P500PN_210790). TJS, MAB, AP, and NP acknowledge support from the Yale Center for Natural Carbon Capture. AK acknowledges funding from Advocates for Climate Innovation. SZ acknowledges support from College of Arts & Sciences Environment and Sustainability Initiative (ESI) at Texas A&M University. NP and CR acknowledge funding from the United States Department of Agriculture (USDA) and the Grantham Foundation for the Environment.

8 Conflict of interest

PAR, CTR, and NJP are scientific advisors to CREW carbon, a company that generates alkalinity in wastewater treatment plants using lime addition but does not work on terrestrial EW.

NJP and CTR were co-founders of the EW supplier Lithos Carbon but have no financial ties to the company.

All other authors declare no conflicts of interest.

9 Data availability statement

All data used for modeling in this study are publicly available as of 22 September 2025 and/or contained within this manuscript. These datasets include: U.S. Geological Survey (USGS) stream gage and river chemistry records (USGS 2016), agricultural land cover data from the National Land Cover Database (NLCD) (USGS 2024), GEOROC database for calculation of US-average basalt composition (Lehnert *et al* 2000), Geochemical and mineralogical data for soils of the conterminous United States to constrain the composition of a large set of agricultural fields (Smith *et al* 2013), data on in-field heterogeneity for 5 field sites (Suhrhoff et al 2025; see supplement S2), constraints on soil CO_2 emissions and variance thereof (Kazula and Lauer 2023, Milliken *et al* 2025).

All analysis and simulation code used in this study is attached to this publication and will be made permanently available via Zenodo upon final publication.

10 Figures

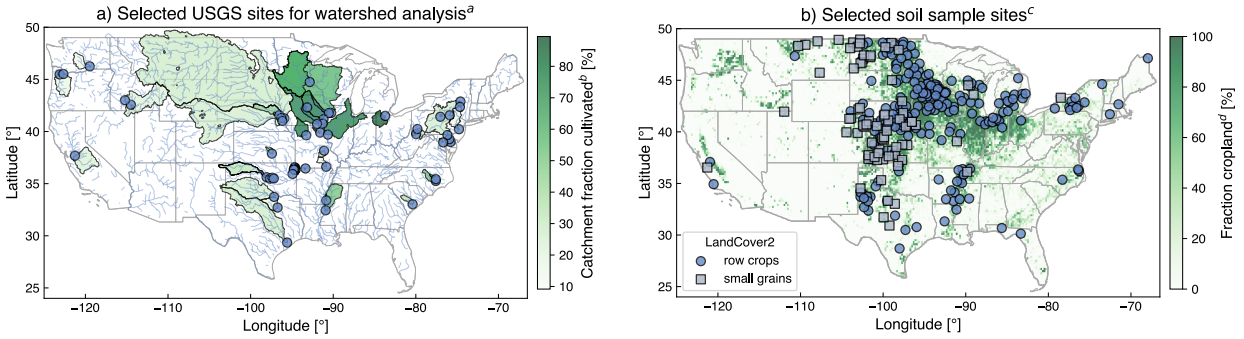

Figure 1: Map of the selected USGS sites for the watershed analysis (a; some watersheds smaller than marker size). Watersheds are color coded by the fraction of land that is classified as cultivated vegetation (Tuanmu and Jetz 2014). Panel (b) shows soil samples used to model aggregated monitoring of soil-based quantification approaches.
[a] (USGS 2016) [b] (Tuanmu and Jetz 2014) [c] (Smith *et al* 2013) [d] (Potapov *et al* 2022)

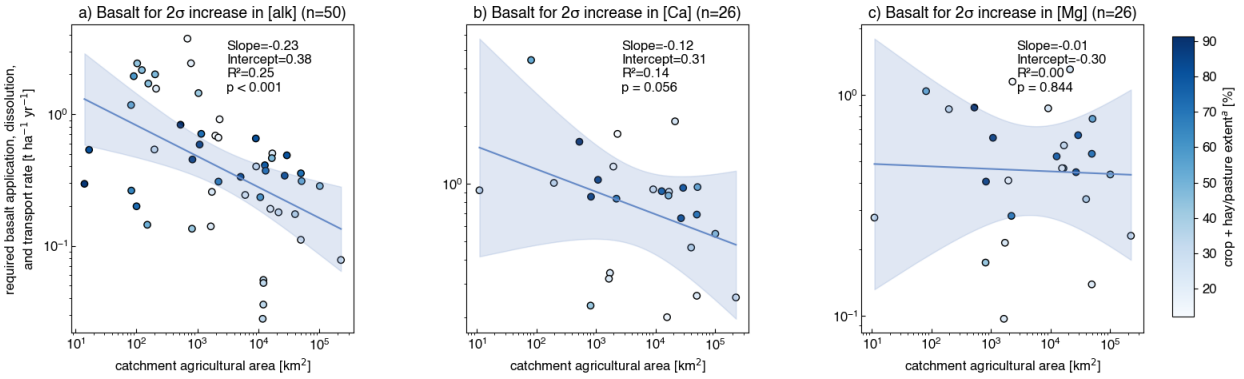

Figure 2: Required agricultural area-normalized basalt dissolution and transport rates to cause a 2σ increase in river alkalinity (a), Ca (b) and Mg (c) concentrations assuming steady state. A version of this Figure but normalized to total catchment area can be found in the supplement (Figure S1)
[a] (USGS 2024)

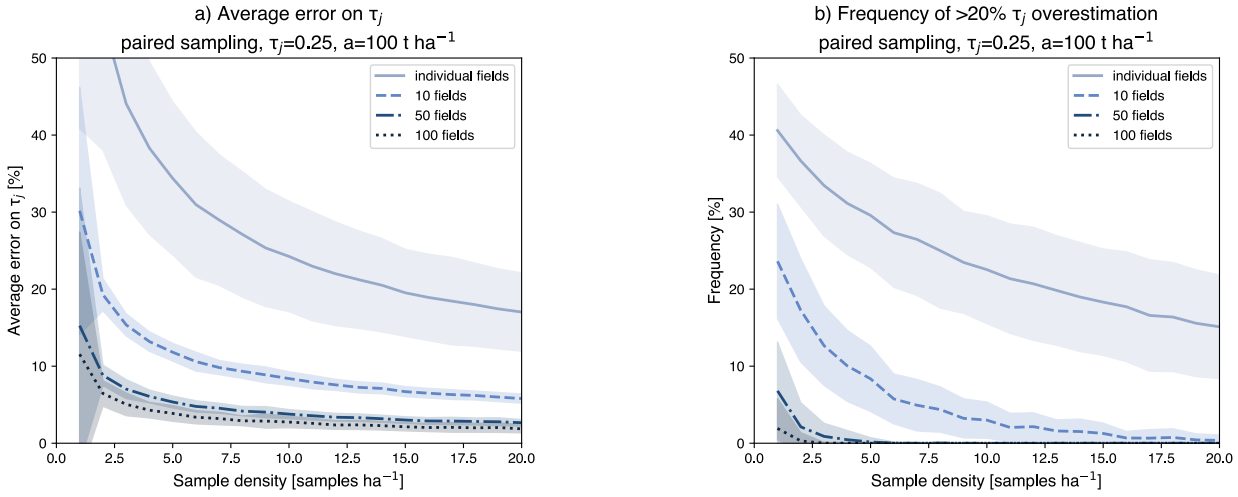

Figure 3: Panel a) shows the error on detected mass transfer coefficients τ_j one would get on average if one applied a soil mass balance approach to quantify rock powder dissolution once for individual as well as sets of fields. The frequency of overestimating τ_j by at least 20% as a function of sampling density and number of aggregated fields is shown in panel b). The plots for the remaining combinations of application amounts and dissolution fractions can be found in the supplement (Figures S3-7). All simulations assume unpaired sampling (see supplement S2.3 for paired sampling).

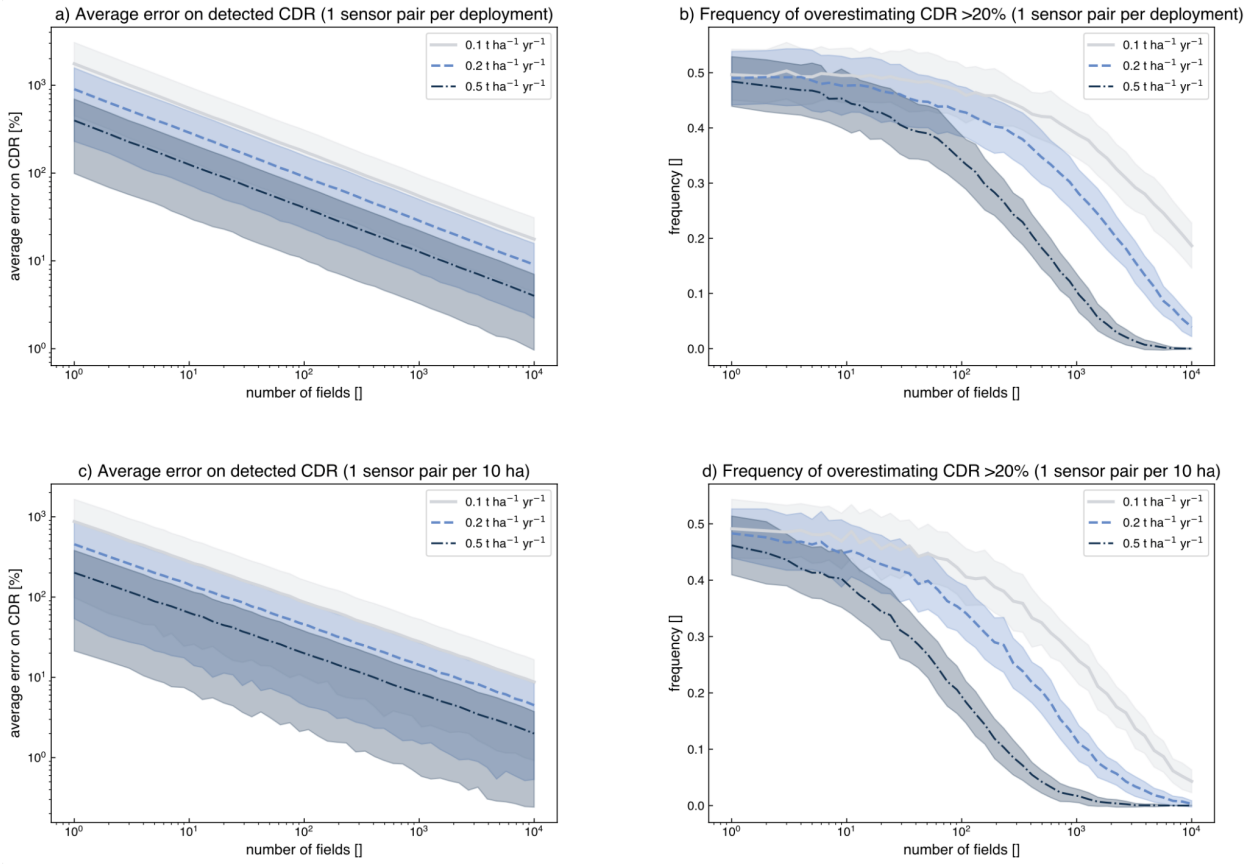

Figure 4: Modeled average error on detected CDR rates using a single pair of control and treatment in-soil CO_2 sensors per deployment as a function of the number of fields being aggregated over as well as nominal CDR rates (a). Panel (b) shows the frequency of overestimating CDR by at least 20% for the same model runs. Panels (c) and (d) simulate the same processes but based on a sensor density of 1 sensor pair per 10 ha.

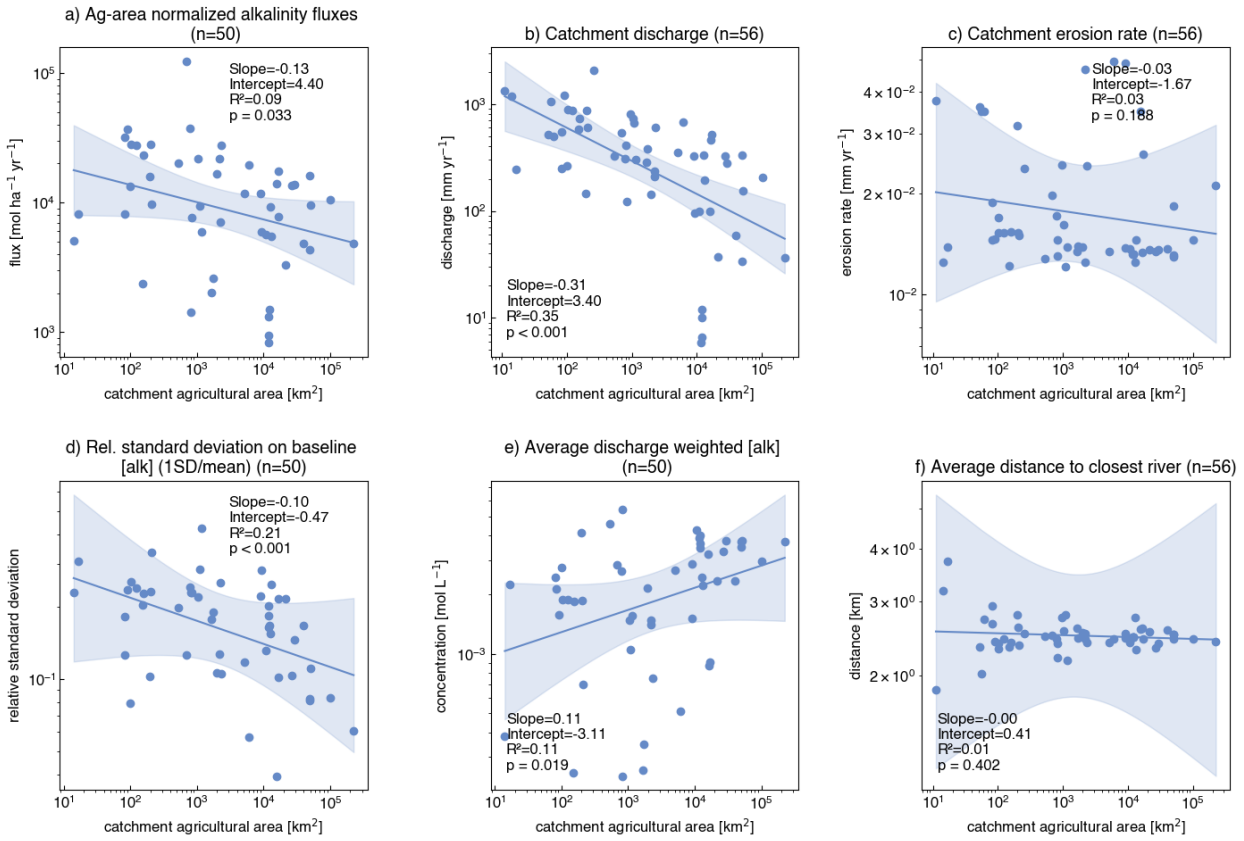

Figure 5: The top row shows agricultural area normalized alkalinity fluxes (a), average annual alkalinity concentration (b), and the variability of baseline alkalinity data (c) for the set of rivers that have sufficient alkalinity baseline data (n=89). These trends are generally similar for Ca and Mg (Figure S3). The bottom row shows area normalized runoff (calculated from discharge and catchment area) (d), average distance to the closest river (e) and erosion rates (f) for all stations that have sufficient baseline data for alkalinity, Ca, or Mg as well as a non-zero proportion of agricultural land (n=56). A version of this figure based on total catchment area, not catchment agricultural area, is included in the supplement (Figure S5).

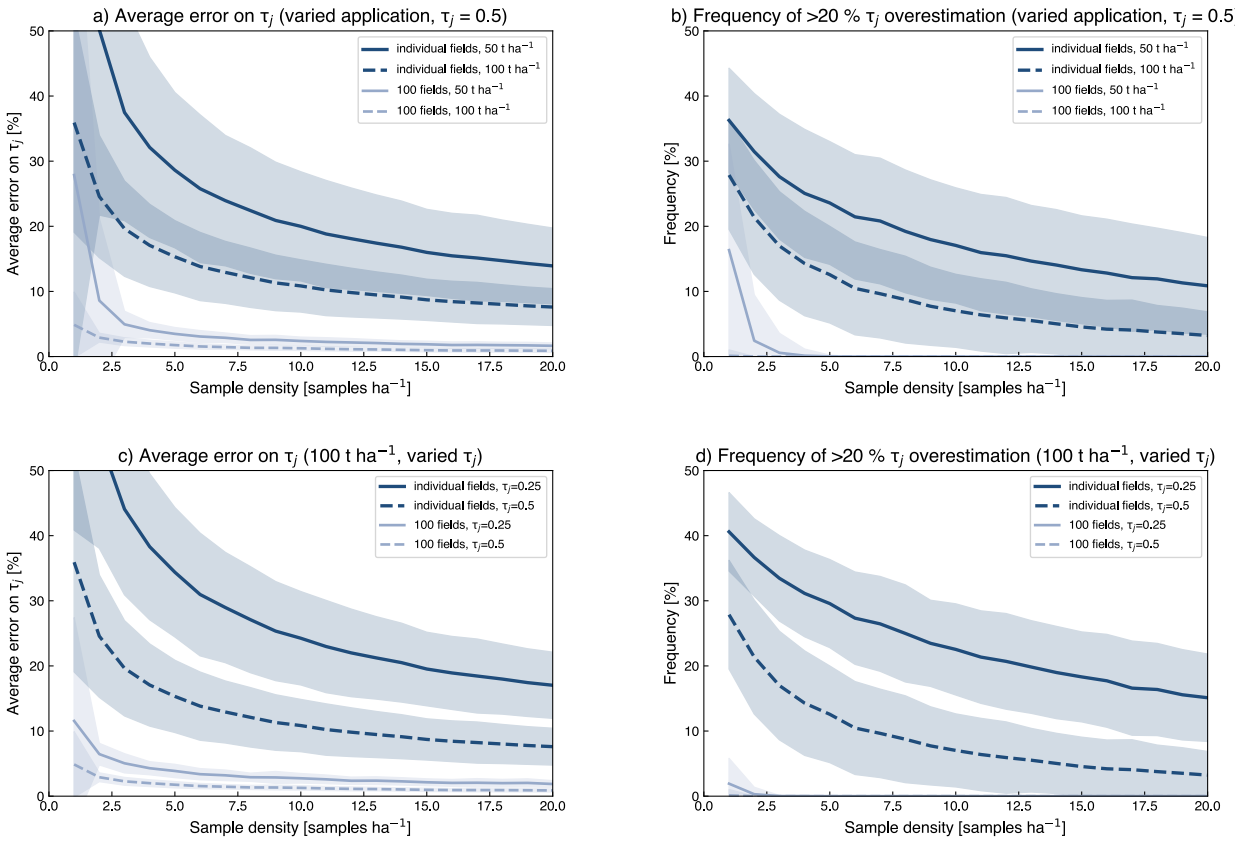

Figure 6: Average error on detected mass transfer coefficients (i.e., dissolution fractions; a and c) as well as the frequency of overestimating mass transfer coefficients by more than 20% (b and d) for constant τ_j but variable application amounts (a and b) as well as at constant application amount but carriable τ_j (c and d). These simulations are based on un-paired sampling (see supplement S2.3 for paired sampling).

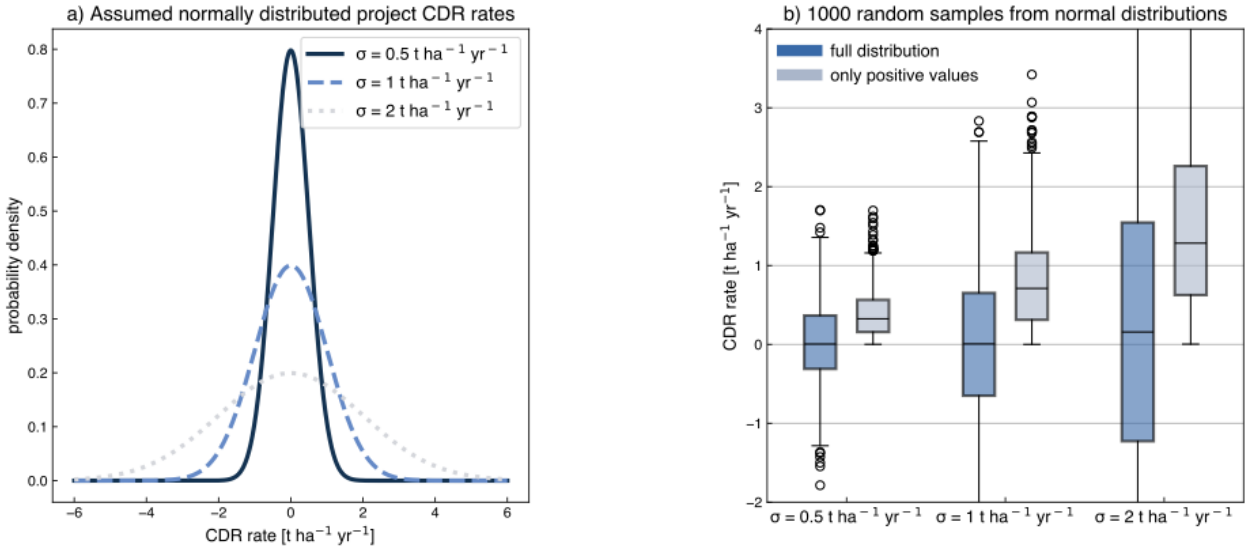

Figure 7: Panel a) shows hypothetical EW deployments where no CDR has occurred. Fields of different heterogeneities (or possibly temporal variability in case of water-based approaches) have varying spread around the mean when samples are used to constrain CDR. If only positive realizations of these random distributions are considered, CDR rates are generated from noise (b). This effect increases the more noise is in the system. Collectively, this demonstrates the necessity to include all fields included in a set when issuing credits as well as that field sets may not be defined based on apparent signal emergence (or exclusion from lack thereof). This effect is not only relevant for no-CDR cases but as long as "negative CDR", i.e., CO2 emissions, is within uncertainty of detected CDR rates.

11 References

Almaraz M, Bingham N L, Holzer I O, Geoghegan E K, Goertzen H, Sohng J and Houlton B Z 2022 Methods for determining the CO2 removal capacity of enhanced weathering in agronomic settings *Frontiers in Climate* **4**

Badgley G, Freeman J, Hamman J J, Haya B, Trugman A T, Anderegg W R L and Cullenward D 2022 Systematic over-crediting in California's forest carbon offsets program *Glob Chang Biol* **28** 1433–45

Baek S H, Kanzaki Y, Lora J M, Planavsky N, Reinhard C T and Zhang S 2023 Impact of Climate on the Global Capacity for Enhanced Rock Weathering on Croplands Earth's Future *Earths Future* **11** 1–12

Barker L J, Fry M, Hannaford J, Nash G, Tanguy M and Swain O 2022 Dynamic High Resolution Hydrological Status Monitoring in Real-Time: The UK Water Resources Portal *Front Environ Sci* **10** 1–15

Barnes R T and Raymond P A 2009 The contribution of agricultural and urban activities to inorganic carbon fluxes within temperate watersheds *Chem Geol* **266** 318–27 Online: http://dx.doi.org/10.1016/j.chemgeo.2009.06.018

Barron-Gafford G A, Scott R L, Jenerette G D and Huxman T E 2011 The relative controls of temperature, soil moisture, and plant functional group on soil CO2 efflux at diel, seasonal, and annual scales *J Geophys Res Biogeosci* **116** 1–16

Beerling D J, Kantzas E P, Lomas M R, Taylor L L, Zhang S, Kanzaki Y, Eufrasio R M, Renforth P, Mecure J F, Pollitt H, Holden P B, Edwards N R, Koh L, Epihov D Z, Wolf A, Hansen J E, Banwart S A, Pidgeon N F, Reinhard C T, Planavsky N J and Val Martin M 2025 Transforming US agriculture for carbon removal with enhanced weathering *Nature* **638** Online: http://dx.doi.org/10.1038/s41586-024-08429-2

Beerling D J, Kantzas E P, Lomas M R, Wade P, Eufrasio R M, Renforth P, Sarkar B, Andrews M G, James R H, Pearce C R, Mercure J, Pollitt H, Holden P B and Edwards N R 2020 Potential for large-scale CO2 removal via enhanced rock weathering with croplands *Nature* **583** 242–8 Online: http://dx.doi.org/10.1038/s41586-020-2448-9

Bijma J, Hagens M, Hammes J, Planavsky N, von Strandmann P A E P, Reershemius T, Reinhard C T, Renforth P, Suhrhoff T J, Vicca S, Vienne A and Wolf-Gladrow D A 2025 Reviews and syntheses: Carbon vs. cation based MRV of Enhanced Rock Weathering and the issue of soil organic carbon *Biogeosciences* 1–29 Online: http://dx.doi.org/10.5194/egusphere-2025-2740

Bradford M A, Eash L, Polussa A, Jevon F V., Kuebbing S E, Hammac W A, Rosenzweig S and Oldfield E E 2023 Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling *Geoderma* **440** 116719 Online: https://doi.org/10.1016/j.geoderma.2023.116719

Bradley V C, Kuriwaki S, Isakov M, Sejdinovic D, Meng X L and Flaxman S 2021 Unrepresentative big surveys significantly overestimated US vaccine uptake *Nature* **600** 695–700

Calabrese S, Wild B, Bertagni M B, Bourg I C, White C, Aburto F, Cipolla G, Noto L V. and Porporato A 2022 Nano- to Global-Scale Uncertainties in Terrestrial Enhanced Weathering *Environ Sci Technol* **56** 15261–72

Campbell J, Bastianini L, Buckman J, Bullock L, Foteinis S, Furey V, Hamilton J, Harrington K, Hawrot O, Holdship P, Knapp W, Maesano C, Mayes W, Pogge von Strandmann P, Reershemius T, Rosair

G, Sturgeon F, Turvey C, Wilson S and Renforth P 2023 *Measurements in Geochemical Carbon Dioxide Removal* (Heriot-Watt University)

CDR.fyi 2025 *Keep Calm and Remove On - The CDR fyi 2024 Year in Review*

Clarkson M O, Larkin C S, Swoboda P, Reershemius T, Suhrhoff T J, Maesano C N and Campbell J S 2024 A Review of Measurement for Quantification of Carbon Dioxide Removal by Enhanced Weathering in Soil *Frontiers in Climate* **6** 1–20

Clausen J, Gerogian T and Bednar A 2013a *Incremental Sampling Methodology (ISM) for Metallic Residues*

Clausen J L, Georgian T and Bednar A 2013b *Cost and Performance Report of Incremental Sampling Methodology for Soil Containing Metallic Residues Engineer Research and Development Center*

Clausen J L, Georgnian T, Bednar A, Perron N, Bray A, Tuminello P, Gooch G, Mulherin N, Gelvin A, Beede M, Saari S, Jones W and Tazik S 2013c *Demonstration of Incremental Sampling Methodology for Soil Containing Metallic Residues*

CRSI 2025 Policy database *Zenodo* Online: http://www.reeep.org/policy-database

Derry L A, Chadwick O A and Porder S 2025 Estimation of Carbon Dioxide Removal via Enhanced Weathering *Glob Chang Biol* **31** 1–3

Dietzen C, Harrison R and Michelsen-Correa S 2018 Effectiveness of enhanced mineral weathering as a carbon sequestration tool and alternative to agricultural lime: An incubation experiment *International Journal of Greenhouse Gas Control* **74** 251–8 Online: https://doi.org/10.1016/j.ijggc.2018.05.007

Duan S, Kaushal S S, Rosenfeldt E J, Murthy S and Fischel M H H 2025 Chemical weathering and land use control river alkalinization and dissolved inorganic carbon in the Potomac River, USA *Applied Geochemistry* **191** 106530 Online: https://doi.org/10.1016/j.apgeochem.2025.106530

Gaillardet J, Dupré B, Louvat P and Allègre C J 1999 Global silicate weathering and CO2 consumption rates deduced from the chemistry of large rivers *Chem Geol* **159** 3–30

Geden O, Gidden M J, Lamb W F, Minx J C, Nemet G F and Smith S M 2024 The State of Carbon Dioxide Removal

Gill-Wiehl A, Kammen D and Haya B 2023 Pervasive over-crediting from cookstoves offset methodologies *Res Sq* 1–23

Gislason S R, Oelkers E H, Eiriksdottir E S, Kardjilov M I, Gisladottir G, Sigfusson B, Snorrason A, Elefsen S, Hardardottir J, Torssander P and Oskarsson N 2009 Direct evidence of the feedback between climate and weathering *Earth Planet Sci Lett* **277** 213–22 Online: http://dx.doi.org/10.1016/j.epsl.2008.10.018

Godsey S E and Kirchner J W 2014 Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order *Hydrol Process* **28** 5791–803

Hadley P W, Crapps E and Hewitt A D 2011 Time for a Change of Scene *Environ Forensics* **12** 312–8

Hamilton S K, Kurzman A L, Arango C, Jin L and Robertson G P 2007 Evidence for carbon sequestration by agricultural liming *Global Biogeochem Cycles* **21** 1–12

Hammond G, Lichtner P and Lu C 2007 Subsurface multiphase flow and multicomponent reactive transport modeling using high-performance computing *J Phys Conf Ser* **78**

Harmel R D, Preisendanz H E, King K W, Busch D, Birgand F and Sahoo D 2023 A Review of Data Quality and Cost Considerations for Water Quality Monitoring at the Field Scale and in Small Watersheds *Water (Switzerland)* **15** 1–19

Hartmann J 2009 Bicarbonate-fluxes and CO2-consumption by chemical weathering on the Japanese Archipelago — Application of a multi-lithological model framework *Chem Geol* **265** 237–71 Online: http://dx.doi.org/10.1016/j.chemgeo.2009.03.024

Heikkinen J, Ketoja E, Nuutinen V and Regina K 2013 Declining trend of carbon in Finnish cropland soils in 1974-2009 *Glob Chang Biol* **19** 1456–69

Hewitt A D, Jenkins T F, Walsh M E, Walsh M R, Bigl S R and Ramsey C A 2007 *Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges for the Characterization of Energetic Munitions Constituents*

Hodges C, Brantley S L, Sharifironizi M, Forsythe B, Tang Q, Carpenter N and Kaye J 2021 Soil Carbon Dioxide Flux Partitioning in a Calcareous Watershed With Agricultural Impacts *J Geophys Res Biogeosci* **126** 1–21

Holland P W 1986 Statistics and causal inference *J Am Stat Assoc* **81** 945–60

Iff N, Renforth P and Pogge von Strandmann P A E 2024 The dissolution of olivine added to soil at 32°C: the fate of weathering products and its implications for enhanced weathering at different temperatures *Frontiers in Climate* **6** 1–18

IPCC 2022 *Climate Change 2022: Mitigation ofClimate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*

IPCC 2018 Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, *Ipcc - Sr15* **2** 17–20 Online: https://www.ipcc.ch/sr15/

IPCC 2024 *IPCC Expert Meeting on Carbon Dioxide Removal Technologies and Carbon Capture Utilization and Storage* ed M Enoki, T., Hayat (Report of the IPCC Expert Meeting, Pub. IGES, Japan) Online: https://www.ipcc-nggip.iges.or.jp/public/mtdocs/pdfiles/2407_Background_CDR_CCUS.pdf

ITRC 2020 *Incremental Sampling Methodology (ISM) Update The Interstate Technology & Regulatory Council (ITRC)*

Kantola I B, Blanc-Betes E, Master M D, Chang E, Marklein A, Moore C E, von Haden A, Bernacchi C J, Wolf A, Epihov D Z, Beerling D J and Delucia E H 2023 Improved net carbon budgets in the US Midwest through direct measured impacts of enhanced weathering *Glob Chang Biol* 1–17

Kanzaki Y, Chiaravalloti I, Zhang S, Planavsky N J and Reinhard C T 2024 In silico calculation of soil pH by SCEPTER v1.0 *Geosci Model Dev* **17** 4515–32

Kanzaki Y, Planavsky N, Zhang S, Jordan J, Suhrhoff T J and Christopher T 2025 Soil cation storage is a key control on the carbon removal dynamics of enhanced weathering *Environmental Research Letters* **20** 1–11

Kanzaki Y, Zhang S, Planavsky N J and Reinhard C T 2022 Soil Cycles of Elements simulator for Predicting TERrestrial regulation of greenhouse gases: SCEPTER v0.9 *Geosci Model Dev* **15** 4959–90

Kazula M J and Lauer J G 2023 Greenhouse gas emissions from soils in corn-based cropping systems *J Environ Qual* **52** 1080–91

Kirchner J W and Neal C 2013 Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. *Proc Natl Acad Sci U S A* **110** 12213–8 Online: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3725102&tool=pmcentrez&rendertype=abstract

Lamb W F, Gasser T, Roman-Cuesta R M, Grassi G, Gidden M J, Powis C M, Geden O, Nemet G, Pratama Y, Riahi K, Smith S M, Steinhauser J, Vaughan N E, Smith H B and Minx J C 2024 The carbon dioxide removal gap *Nat Clim Chang* **14** 644–51 Online: http://dx.doi.org/10.1038/s41558-024-01984-6

Lehnert K, Su Y, Langmuir C H, Sarbas B and Nohl U 2000 A global geochemical database structure for rocks *Geochemistry, Geophysics, Geosystems* **1**

Levy C R, Almaraz M, Beerling D J, Raymond P, Reinhard C T, Suhrhoff T J and Taylor L 2024 Enhanced Rock Weathering for Carbon Removal − Monitoring and Mitigating Potential Environmental Impacts on Agricultural Land *Environ Sci Technol* **58** 17215–26

Luderer G, Vrontisi Z, Bertram C, Edelenbosch O Y, Pietzcker R C, Rogelj J, De Boer H S, Drouet L, Emmerling J, Fricko O, Fujimori S, Havlík P, Iyer G, Keramidas K, Kitous A, Pehl M, Krey V, Riahi K, Saveyn B, Tavoni M, Van Vuuren D P and Kriegler E 2018 Residual fossil CO2 emissions in 1.5-2 °c pathways *Nat Clim Chang* **8** 626–33 Online: http://dx.doi.org/10.1038/s41558-018-0198-6

Maier M and Schack-Kirchner H 2014 Using the gradient method to determine soil gas flux: A review *Agric For Meteorol* **192–193** 78–95 Online: http://dx.doi.org/10.1016/j.agrformet.2014.03.006

Maillard É, McConkey B G and Angers D A 2017 Increased uncertainty in soil carbon stock measurement with spatial scale and sampling profile depth in world grasslands: A systematic analysis *Agric Ecosyst Environ* **236** 268–76 Online: http://dx.doi.org/10.1016/j.agee.2016.11.024

Mercer L, Burke J and Rodway-dyer S 2024 *Towards improved cost estimates for monitoring , reporting and verification of carbon dioxide removal* (London School of Economics and Political Science)

Milliken E, Woodley A and Planavsky N J 2025 Direct Measurement of Carbon Dioxide Removal Due to Enhanced Weathering *Environ Sci Technol Lett*

Mills R T, Lu C, Lichtner P C and Hammond G E 2007 Simulating subsurface flow and transport on ultrascale computers using PFLOTRAN *J Phys Conf Ser* **78**

Moosdorf N, Hartmann J, Lauerwald R, Hagedorn B and Kempe S 2011 Atmospheric CO2 consumption by chemical weathering in North America *Geochim Cosmochim Acta* **75** 7829–54 Online: http://dx.doi.org/10.1016/j.gca.2011.10.007

NASEM 2020 *Review of the New York City Watershed Protection Program* (National Academies of Sciences, Engineering, and Medicine;The National Academies Press, Washington, DC)

Oh N H and Raymond P A 2006 Contribution of agricultural liming to riverine bicarbonate export and CO2 sequestration in the Ohio River basin *Global Biogeochem Cycles* **20** 1–17

Oldfield B E E, Eagle A J, Rubin R L, Rudek J and Gordon D R 2022 Crediting agricultural soil carbon sequestration; Regional consistency is necessary for carbon credit integrity *Science (1979)* **375** 1222–5

te Pas E E E M, Chang E, Marklein A R, Comans R N J and Hagens M 2025 Accounting for retarded weathering products in comparing methods for quantifying carbon dioxide removal in a short-term enhanced weathering study *Frontiers in Climate* **1524998** 1–10

Poeplau C and Don A 2015 Carbon sequestration in agricultural soils via cultivation of cover crops - A meta-analysis *Agric Ecosyst Environ* **200** 33–41 Online: http://dx.doi.org/10.1016/j.agee.2014.10.024

Pogge von Strandmann P A E, Tooley C, Mulders J J P A and Renforth P 2022 The Dissolution of Olivine Added to Soil at 4°C: Implications for Enhanced Weathering in Cold Regions *Frontiers in Climate* **4** 1–11

Potapov P, Turubanova S, Hansen M C, Tyukavina A, Zalles V, Khan A, Song X P, Pickens A, Shen Q and Cortez J 2022 Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century *Nat Food* **3** 19–28

Potash E, Bradford M A, Oldfield E E and Guan K 2025 Measure-and-remeasure as an economically feasible approach to crediting soil organic carbon at scale *Environmental Research Letters* **20**

Power I M, Hatten V N J, Guo M, Rausis K and Klyn-hesselink H 2025 Are enhanced rock weathering rates overestimated? A few geochemical and mineralogical pitfalls 1–9

Puro.Earth 2024 Enhanced Rock Weathering Methodology for CO2 Removal 1–68 Online: https://7518557.fs1.hubspotusercontent-na1.net/hubfs/7518557/Supplier Documents/ERW methodology.pdf

Rausis K, Stubbs A R, Power I M and Paulo C 2022 Rates of atmospheric CO2 capture using magnesium oxide powder *International Journal of Greenhouse Gas Control* **119** 103701 Online: https://doi.org/10.1016/j.ijggc.2022.103701

Reershemius T, Kelland M E, Davis I R, D'Ascanio R, Kalderon-Asael B, Asael D, Suhrhoff T J, Epihov D E, Beerling D J, Reinhard C T and Planavsky N J 2023 Initial Validation of a Soil-Based Mass-Balance Approach for Empirical Monitoring of Enhanced Rock Weathering Rates *Environ Sci Technol* **57** 19497–19507 Online: http://arxiv.org/abs/2302.05004

Reershemius T and Suhrhoff T J 2023 On error, uncertainty, and assumptions in calculating carbon dioxide removal rates by enhanced rock weathering in Kantola et al ., 2023 *Glob Chang Biol* 1–3

Renforth P and Henderson G 2017 Assessing ocean alkalinity for carbon sequestration *Reviews of Geophysics* **55** 636–74

Renforth P, Pogge von Strandmann P A E and Henderson G M 2015 The dissolution of olivine added to soil: Implications for enhanced weathering *Applied Geochemistry* **61** 109–18 Online: http://dx.doi.org/10.1016/j.apgeochem.2015.05.016

Rogelj J, Popp A, Calvin K V., Luderer G, Emmerling J, Gernaat D, Fujimori S, Strefler J, Hasegawa T, Marangoni G, Krey V, Kriegler E, Riahi K, Van Vuuren D P, Doelman J, Drouet L, Edmonds J, Fricko O, Harmsen M, Havlík P, Humpenöder F, Stehfest E and Tavoni M 2018 Scenarios towards limiting global mean temperature increase below 1.5 °c *Nat Clim Chang* **8** 325–32 Online: http://dx.doi.org/10.1038/s41558-018-0091-3

Rogers B and Maher K 2025 A Framework for Integrating Spatial Uncertainty into Critical Zone Models : Application to Enhanced Weathering *CDRXIV*

Rothman K J, Greenland S and Lash T L 2008 *Modern epidemiology* vol 3 (Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia)

Rubin D B 1974 Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* **66** 688

Sanders-DeMott R, Hutyra L R, Hurteau M D, Keeton W S, Fallon K S, Anderegg W R L, Hollinger D Y, Kuebbing S E, Lucash M S, Ordway E M, Vargas R and Walker W S 2025 Ground-Truth: Can Forest Carbon Protocols Ensure High-Quality Credits? *Earths Future* **13**

Shao S, Driscoll C T, Johnson C E, Fahey T J, Battles J J and Blum J D 2016 Long-term responses in soil solution and stream-water chemistry at Hubbard Brook after experimental addition of wollastonite *Environmental Chemistry* **13** 528–40

Shaughnessy A R and Brantley S L 2023 How do silicate weathering rates in shales respond to climate and erosion? *Chem Geol* **629**

Shaughnessy A R, Forgeng M J, Wen T, Gu X, Hemingway J D and Brantley S L 2023 Linking Stream Chemistry to Subsurface Redox Architecture *Water Resour Res* **59** 1–20

Shoch D, Swails E, Sonne Hall E, Belair E, Rifkin B, Griscom B and Latta G 2024 VM0045 - Improved Forest Management Using Dynamic Matched Baselines From National Forest Inventories 1–70

Smith D B, Cannon W F, Woodruff L G, Solano F, Kilburn J E and Fey D L 2013 Geochemical and mineralogical data for soils of the conterminous United States *U.S. Geological Survey Data Series* **801** 1–26

Steefel C I and Molins S 2009 CrunchFlow *Software for modeling multicomponent reactive flow and transport. User's manual* 12–91

Stubbs A R, Paulo C, Power I M, Wang B, Zeyen N and Wilson S A 2022 Direct measurement of CO2 drawdown in mine wastes and rock powders: Implications for enhanced rock weathering *International Journal of Greenhouse Gas Control* **113** 103554 Online: https://doi.org/10.1016/j.ijggc.2021.103554

Suhrhoff T J, Reershemius T, Jordan J, Li S, Zhang S, Kalderon-asael B, Ebert Y, Nyateka R, Reinhard C T, Planavsky N J, Haven N, Sciences P, Haven N, Carbon M, Sciences A and Kingdom U 2025 Updated framework and signal-to-noise analysis of soil mass balance approaches for quantifying enhanced weathering on managed lands *CDRXIV* 1–59

Suhrhoff T J, Reershemius T, Wang J, Jordan J S, Reinhard C T and Planavsky N J 2024 A tool for assessing the sensitivity of soil-based approaches for quantifying enhanced weathering: a US case study *Frontiers in Climate* **6** 1–17

Susha I, Rukanova B, Zuiderwijk A, Gil-Garcia J R and Gasco Hernandez M 2023 Achieving voluntary data sharing in cross sector partnerships: Three partnership models *Information and Organization* **33** 100448 Online: https://doi.org/10.1016/j.infoandorg.2023.100448

Sutherland K, Holme E, Savage R, Gill S, Matlin-Wainer M, He J, Marsland E and Patel C 2024 Isometric Enhanced Weathering in Agriculture v1.0 Online: https://registry.isometric.com/protocol/enhanced-weathering-agriculture

Taylor L L, Beerling D J, Quegan S and Banwart S A 2017 Simulating carbon capture by enhanced weathering with croplands: An overview of key processes highlighting areas of future model development *Biol Lett* **13**

Taylor L L, Driscoll C T, Groffman P M, Rau G H, Blum J D and Beerling D J 2021 Increased carbon capture by a silicate-treated forested watershed affected by acid deposition *Biogeosciences* **18** 169–88

Tuanmu M N and Jetz W 2014 A global 1-km consensus land-cover product for biodiversity and ecosystem modelling *Global Ecology and Biogeography* **23** 1031–45

UNEP 2024 *Emissions Gap Report 2024: No more hot air … please! With a massive gap between rhetoric and reality, countries draft new climate commitments* (United Nations Environment Program, Nairobi, Kenya)

USDA 2024 Farms and Farmland *2022 Census of Agriculture Highlights*

USDA 2022 Farms and Land in Farms 2021 Summary *United States Department of Agriculture: National Agricultural Statistics Service* 1995–2004 Online: https://www.nass.usda.gov/Publications/Todays_Reports/reports/fnlo0222.pdf

USGS 2023 CroplandCROS 2022 Online: https://pdi.scinet.usda.gov/portal/apps/sites/#/cropcros/pages/download-data

USGS 2024 *National Land Cover Database (NLCD) 2021 Land Cover (CONUS)* Online: https://www.mrlc.gov/data

USGS 2016 National Water Information System data available on the World Wide Web (USGS Water Data for the Nation) Online: http://waterdata.usgs.gov/nwis/

Vienne A, Frings P, Poblador S, Steinwidder L, Rijnders J, Schoelynck J, Vinduskova O and Vicca S 2024 Earthworms in an enhanced weathering mesocosm experiment: effects on soil carbon sequestration, base cation exchange and soil CO2 efflux *Soil Biol Biochem* **199** 109596 Online: https://doi.org/10.1016/j.soilbio.2024.109596

Vienne A, Frings P, Rijnders J, Suhrhoff T J, Reershemius T, Poetra R P, Hartmann J, Niron H, Estrada M P, Steinwidder L, Boito L and Vicca S 2025 Weathering without inorganic CDR revelead through cation tracing *EGUsphere*

West J A 2012 Thickness of the chemical weathering zone and implications for erosional and climatic drivers of weathering and for carbon-cycle feedbacks *Geology* **40** 811–4

White A F and Blum A E 1995 Effects of climate on chemical weathering in watersheds *Geochim Cosmochim Acta* **59** 1729–47

Wieczorek M E 2019 Area- and Depth- Weighted Average of Soil pH from STATSGO2 for the Conterminous United States and District of Columbia

Woollen B J and Planavsky N J 2024 Common-Pool Resource Management Problems in Enhanced Weathering Deployments for Carbon Dioxide Removal *Workshop on the Ostrom Workshop – WOW7 Conference* Online: https://hdl.handle.net/10535/10968

Yan Y, Dong X, Li R, Zhang Y, Yan S, Guan X, Yang Q, Chen L, Fang Y, Zhang W and Wang S 2023 Wollastonite addition stimulates soil organic carbon mineralization: Evidences from 12 land-use types in subtropical China *Catena (Amst)* **225** 107031 Online: https://doi.org/10.1016/j.catena.2023.107031

Zhang S, Reinhard C T, Liu S, Kanzaki Y and Planavsky N J 2025 A framework for modeling carbon loss from rivers following terrestrial enhanced weathering *Environmental Research Letters* **20**

Supplementary information to "Aggregated monitoring of enhanced weathering on agricultural lands"

Tim Jesper Suhrhoff[1,2], Anu Khan[3], Shuang Zhang[4], Beck Woollen[1,3], Tom Reershemius[5], Mark A. Bradford[1,6], Alexander Polussa[1,6], Ella Milliken[2], Pete Raymond[6], Christopher T. Reinhard[7], Noah J. Planavsky[2,1]

1: Yale Center for Natural Carbon Capture, New Haven, CT, USA
2: Department of Earth & Planetary Sciences, Yale University, New Haven, CT, USA
3: Carbon Removal Standards Initiative, Washington D.C., USA
4: Department of Oceanography, Texas A&M University, College Station, TX, USA
5: School of Natural and Environmental Sciences, Newcastle University, Newcastle-upon-Tyne, UK
6: The Forest School, Yale School of the Environment, Yale University, New Haven, CT, USA
7: Department of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

S1 Supplementary Figures

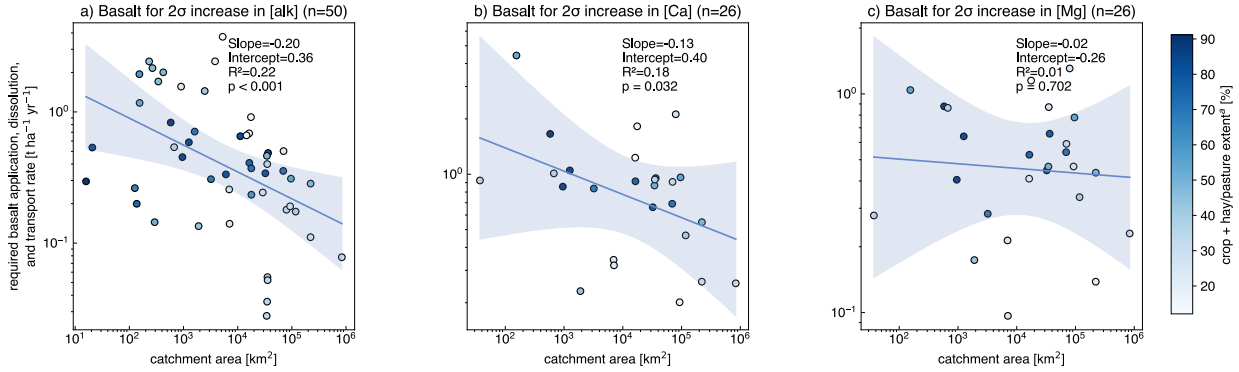

Figure S1: Required total catchment area normalized basalt dissolution and transport rates to cause a 2σ increase in river alkalinity (a), Ca (b) and Mg (c) concentrations assuming steady state.

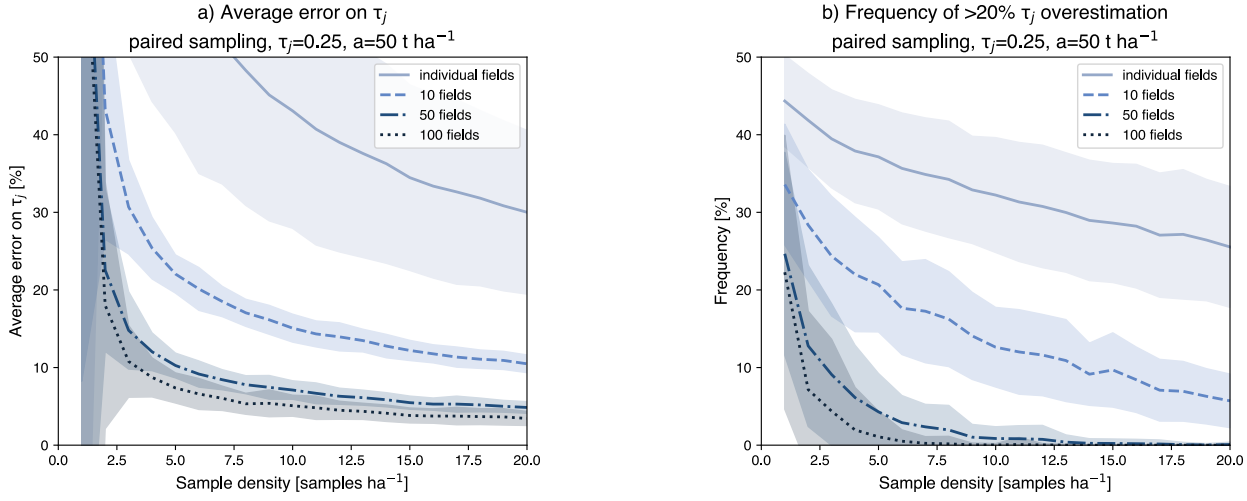

Figure S2: Equivalent to Figure 3 but for τ_j = 0.25 and a = 50 t ha^{-1}. Panel a) shows the error on detected mass transfer coefficients τ_j one would get on average if one applied a soil mass balance approach to quantify rock powder dissolution once for individual as well as sets of fields. The frequency of overestimating τ_j by at least 20% as a function of sampling density and number of aggregated fields is shown in panel b). The plots for the remaining combinations of application amounts and dissolution fractions can be found in the supplement (Figures S3-7). All simulations assume unpaired sampling (see supplement S2.3 for paired sampling).

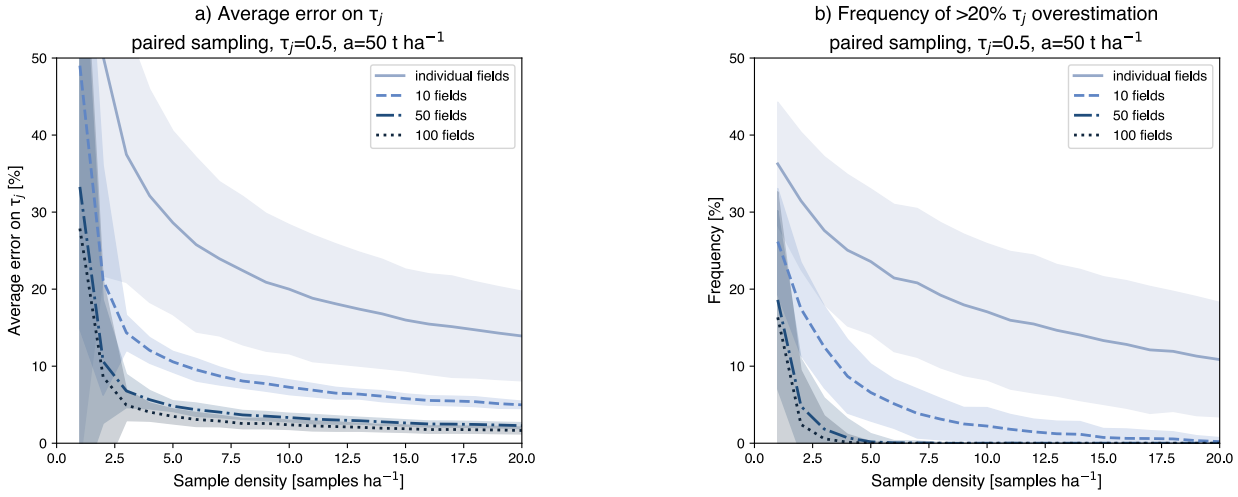

Figure S3: Equivalent to Figure 3 but for τ_j = 0.5 and a = 50 t ha^{-1}. Panel a) shows the error on detected mass transfer coefficients τ_j one would get on average if one applied a soil mass balance approach to quantify rock powder dissolution once for individual as well as sets of fields. The frequency of overestimating τ_j by at least 20% as a function of sampling density and number of aggregated fields is shown in panel b). The plots for the remaining combinations of application amounts and dissolution fractions can be found in the supplement (Figures S3-7). All simulations assume unpaired sampling (see supplement S2.3 for paired sampling).

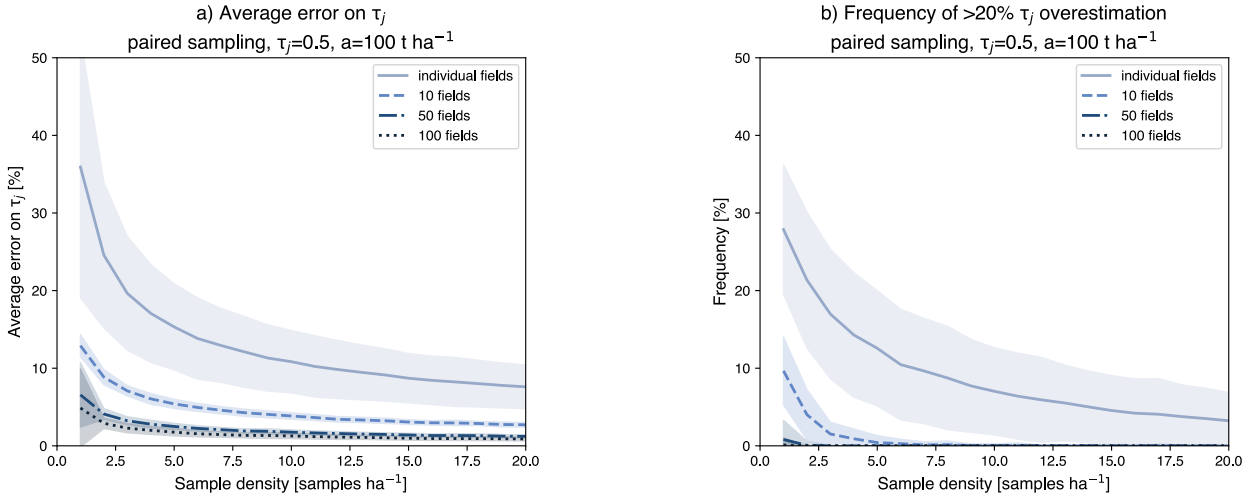

Figure S4: Equivalent to Figure 3 but for τ_j = 0.25 and a = 50 t ha^{-1}. Panel a) shows the error on detected mass transfer coefficients τ_j one would get on average if one applied a soil mass balance approach to quantify rock powder dissolution once for individual as well as sets of fields. The frequency of overestimating τ_j by at least 20% as a function of sampling density and number of aggregated fields is shown in panel b). The plots for the remaining combinations of application amounts and dissolution fractions can be found in the supplement (Figures S3-7). All simulations assume unpaired sampling (see supplement S2.3 for paired sampling).

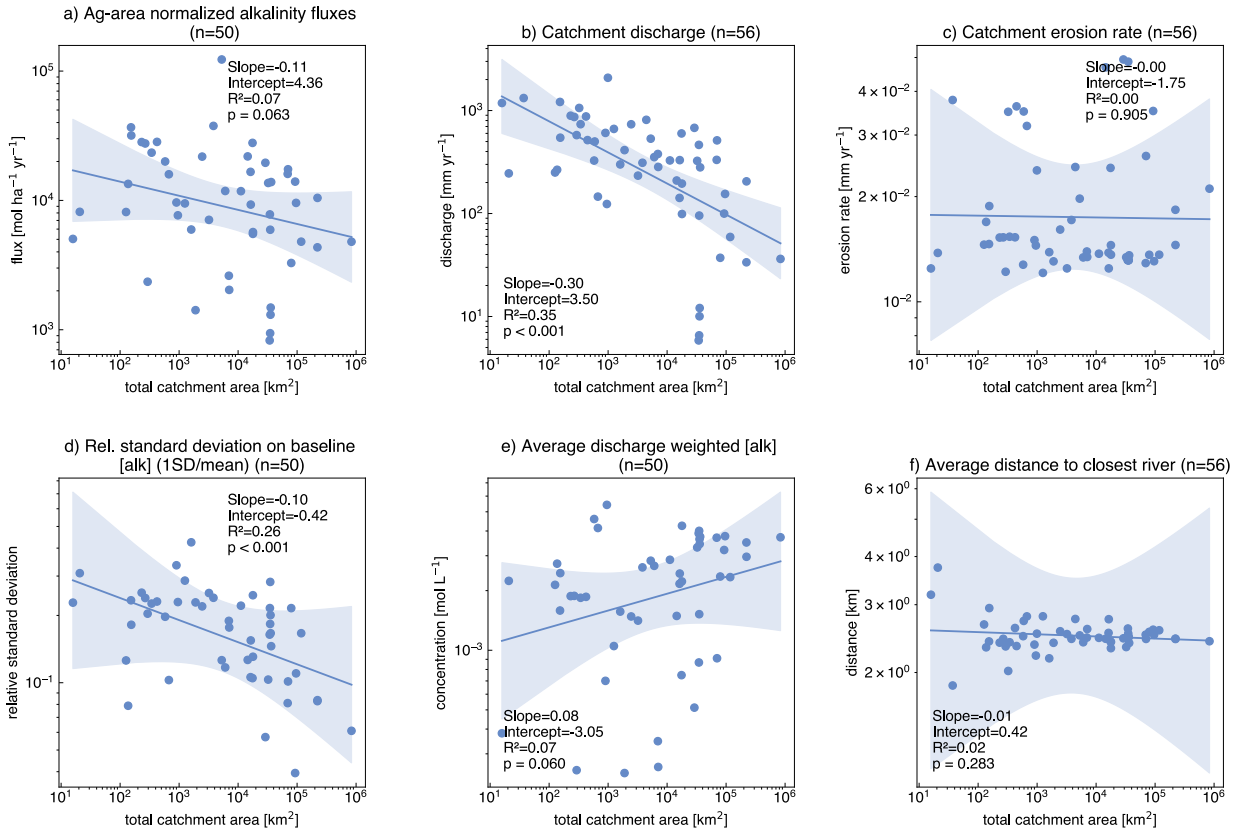

Figure S5: The top row shows total catchment area normalized alkalinity fluxes (a), average annual alkalinity concentration (b), and the variability of baseline alkalinity data (c) for the set of rivers that have sufficient alkalinity baseline data (n=89). The bottom row shows area normalized runoff (d), average distance to the closest river (e) and erosion rates (f) for all stations that have sufficient baseline data for alkalinity, Ca, or Mg as well as a non-zero proportion of agricultural land (n=118). The equivalent information of panels a, d, and e for Ca and Mg can be found in Figure S6.

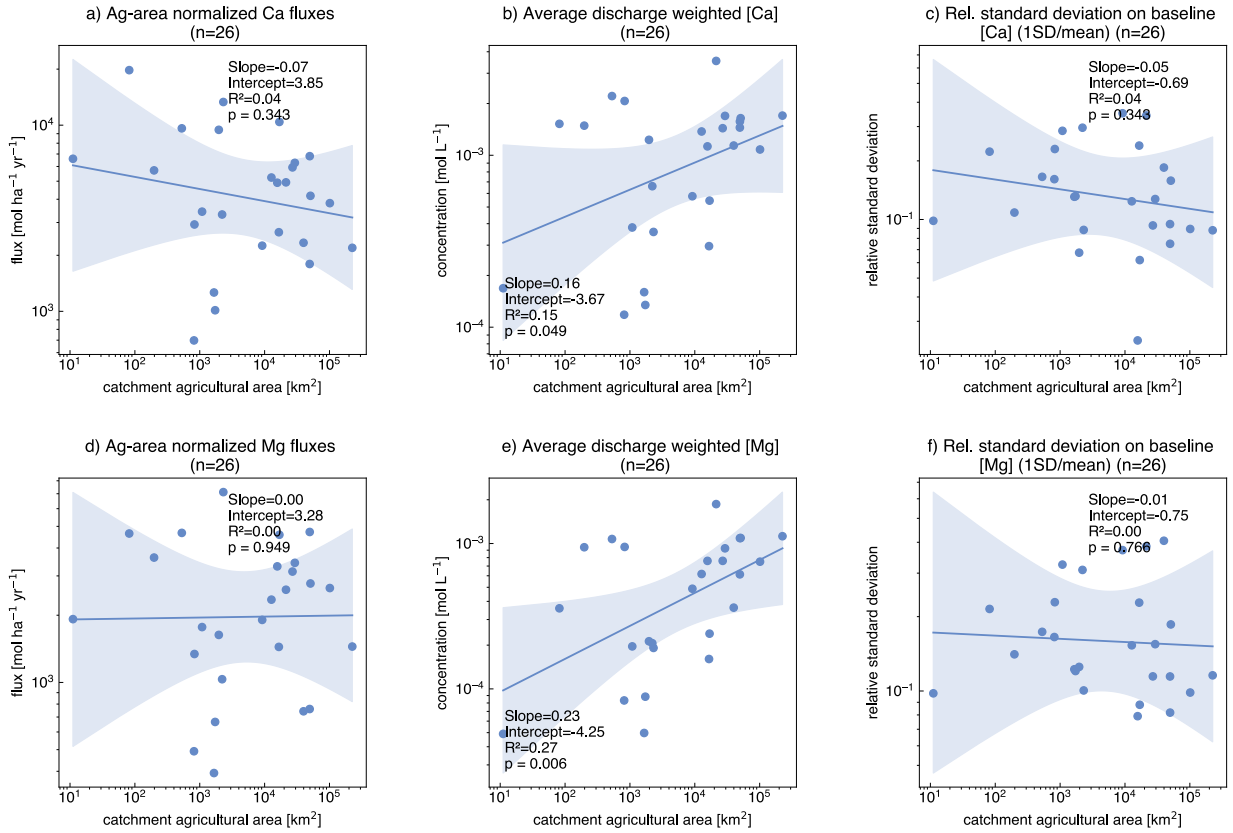

Figure S6: The top row shows agricultural area normalized Ca fluxes (a), average annual Ca concentration (b), and the variability of baseline Ca data (c). The bottom row shows the same type of data but for Mg (d-f).

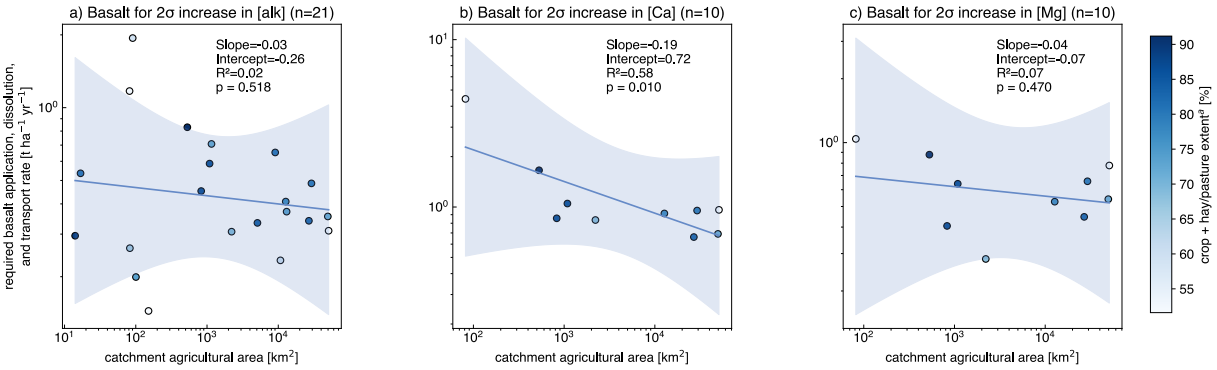

Figure S7: Required basalt application rates to cause a 2σ increase in river alkalinity (a), Ca (b) and Mg (c) concentrations in catchments where more than 50% of the area is classified as crop or hay/pasture land.

[a](USGS, 2024)

S2 Soil signal-to-noise Monte Carlo simulations

S2.1 Method details

To complement the simplified description of our soil analysis framework presented in the main text, we provide here a more detailed account of the data sources, assumptions, and modeling steps underlying the Monte Carlo simulations. This expanded description also outlines how baseline and post-weathering sample compositions are generated, how heterogeneity is parameterized, and how error metrics are derived for both single-field and aggregated-field applications.

The analysis uses multiple data sources to constrain the elemental composition of agricultural fields, rock powder, and expected in-field heterogeneity. To ground this analysis in realistic data of soil composition, we use US soil data classified as "Row Crops" and "Small Grains" (LandCover2) within the "Geochemical and mineralogical data for soils of the conterminous United States" database (Smith *et al* 2013). These samples are treated as the "true" baseline composition of fields, each datapoint in the database being used as one representative field composition. The composition for rock powder is based on the average composition of all samples contained in the GEOROC database that are classified as basalt and contained within the conterminous US (Lehnert *et al* 2000). Because the framework only works for feedstock-soil combinations whose composition is sufficiently different (Suhrhoff *et al* 2024), we only consider soil compositions where both base cation (here Ca^{2+} and Mg^{2+}) as well as immobile element (Ti) concentrations are at least 4 times lower than for basalt (n=302; see Figure *1*b).

For each field, we calculate a "true" post-weathering soil-feedstock mix composition based on assumed application amounts (a= 50 and 100 t ha^{-1}) and dissolution fractions (τ_j = 0.25 and 0.5) (Suhrhoff *et al* 2025). Note that application amount corresponds to the total cumulative amount deployed. Many EW studies apply 40 t ha^{-1} yr^{-1} such that even the highest rate modeled here may be realistic after multiple years of deployments (Beerling *et al* 2020, 2025). Furthermore, for each field a size between 10 and 100 ha is randomly generated (uniform distributions), which is a compromise between skewed US farm size distributions with most farm land being in farms larger than 2,000 ha but most farms being smaller than 72 ha (USDA 2022, 2024).

To constrain variance on field-level sample compositions resulting from spatial heterogeneity, we use a new dataset based on high-density spatial sampling (Suhrhoff et al 2025; cf. S2.2 below). This dataset of soil heterogeneity is based on new ICP-MS soil composition measurements (residual phase after exchangeable cations were leached with 1M ammonium acetate) from 5 field sites in the US with spatial sampling frequencies ranging from 0.6 – 19.8 samples ha^{-1} (7.1 – 39.6 pooled sub samples ha^{-1}). We fit log-normal distributions to field data (using the Python scipy.stats module), and use fitted shape parameters representing the standard deviations (σ) of the underlying normal distribution to model in-field variance– see section S2.2 for more detail on log-normal fits to field data. The shape parameters corresponding to field data are shown in Figure S9 and Figure S10, and uniform distributions between the range of observed shape parameters is used to generate synthetic σ values for Monte Carlo simulations of baseline soils where the resulting distributions are scaled such that the mean of the log normal distribution is equivalent to the field mean (see supplement S2.2). For feedstocks, heterogeneity is introduced by generating shape parameters of 5–10% (uniform distribution, i.e. σ values of 0.05 to 0.1 for generated log normal distributions).

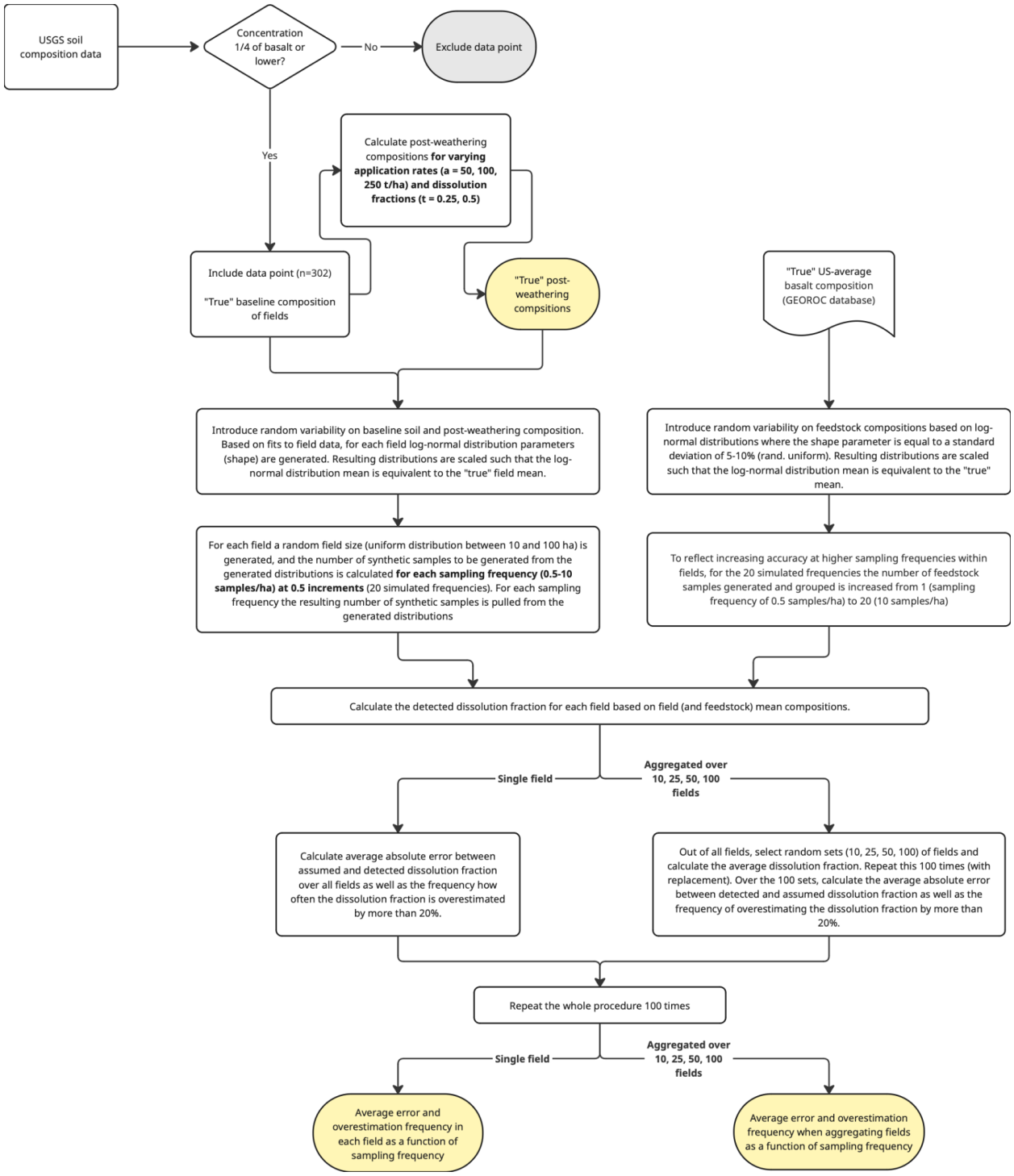

Figure S8: Flow chart of the Monte Carlo simulation algorithm used for the main analysis. This approach assumes non-paired sampling

We use these data to generate a simple statistical model based on nested Monte Carlo simulations (see Figure S8) to assess expected errors on calculated dissolution fractions. The simulated in-field heterogeneity is used to generate baseline and post-weathering soil sample compositions for a range of sampling frequencies. To reflect increasing thoroughness of the sampling approach, as soil sampling

frequency increases from 1 to 20 samples ha^{-1} we also increase the number of total samples that the composition of the feedstock endmember is calculated from (from 1 to 20 samples). Next, the average composition of the realized samples is used to calculate the corresponding dissolution fraction. This is compared to the assumed, true value. This is done for each field 100 times, and we 1) calculate the average error over all fields for each sampling frequency as a realistic error estimate given US soil composition and realistic spatial heterogeneity, and 2) based on first averaging the calculated dissolution fraction based on realized sample composition over multiple fields (10, 25, and 50, done 100 fields based on random pulls from the 302 fields; 100 times) before computing the average error on the dissolution fraction (see Figure S8 for details). These reflect the error that one would get if applying such soil-based mass balance framework as the basis to quantifying CDR on average if applying it once either to an individual field or an aggregated set of fields at the same time. We use these error rates to assess the frequency with which rock powder dissolution is overestimated by 20% or more.

While the dissolution fraction is only a proxy for CDR, this approach can be translated to CDR as well using generated field sizes, application amounts, dissolution fractions, and assumed loss fractions, but we focus here on the dissolution fraction (or mass transfer coefficient) as it is the primary measured quantity. Note that by independently generating baseline and post-weathering samples, we model a non-paired sampling approach reflecting the worst case scenario where paired sampling is either not attempted or prevented due to bad GPS accuracy (typical GPS accuracy of 5 m). We also include additional analysis of paired sampling (cf. S2.3; Figure S11 and Figure S12).

S2.2 Implementation of soil heterogeneity in Monte Carlo simulations

We use soil composition data from five novel field sites sampled at high spatial densities to constrain in-field heterogeneity for the Monte Carlo signal-to-noise analysis. The data are normalized by the field mean concentration (Figure S9) before we fit log-normal distributions to make sure the population means are 1. The use of log-normal (rather than normal) distributions is intentional because samples generated from log-normal distributions always have positive values, preventing the occurrence of non-physical negative soil concentrations in the signal-to-noise analysis without having to filter some data. For normal distributions, this could be achieved by simply filtering out negative model occurrences, but this would change the mean of generated sample distributions and cause a systematic error in calculated dissolution fractions. In addition, using log-normal compared to normal fits also represents a conservative choice for the signal-to-noise analysis due to the generally higher variance, as well has overall better fits compared to normal distributions (R^2 better for 11 out of 20 elemental field distributions).

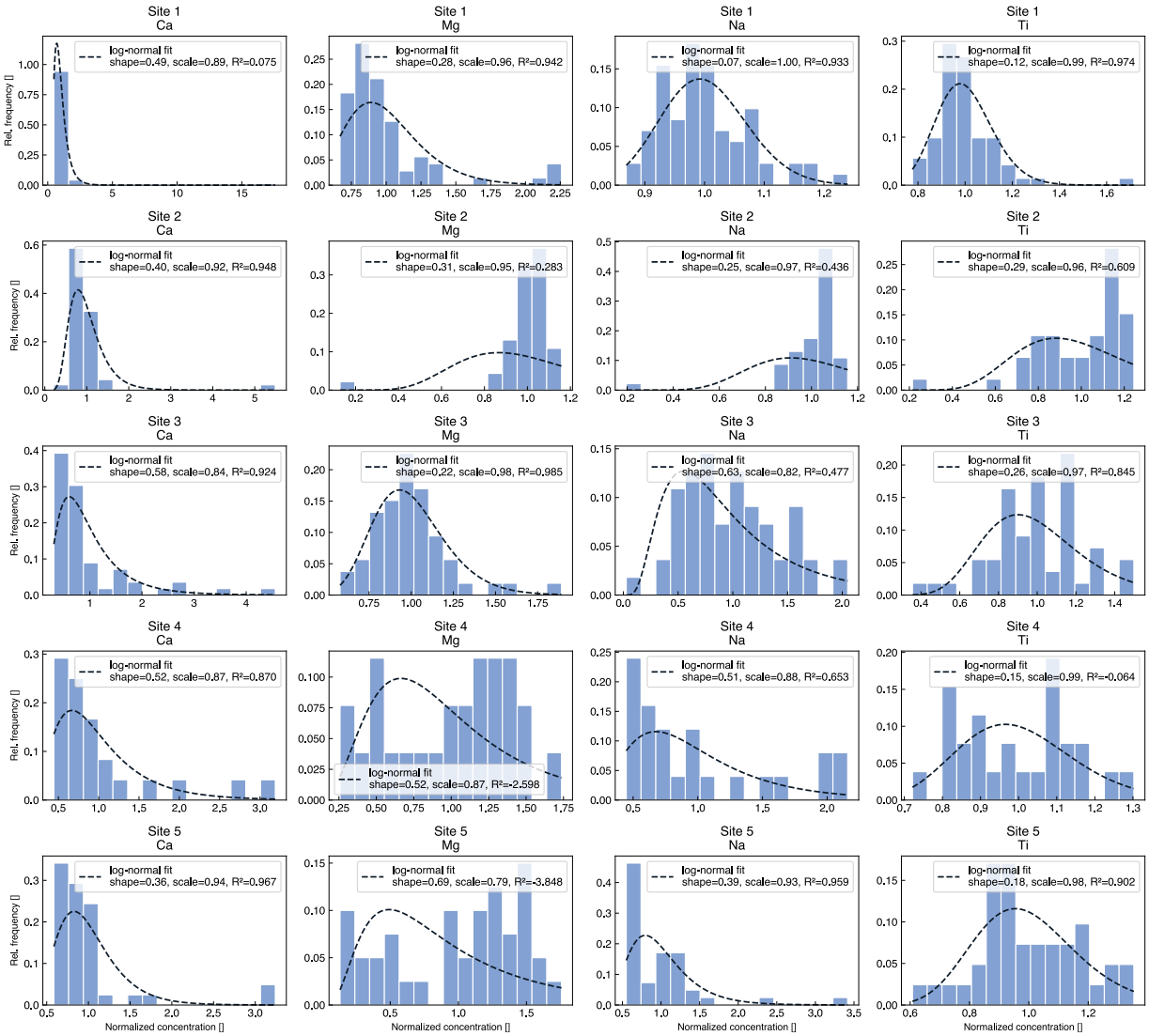

Figure S9: Distributions of baseline data for the 5 field sites (see Table S1 and Suhrhoff et al (2025) for more details) including log-normal fits to the data. The shape parameters, corresponding to the standard deviation of the normal distribution of the logarithm of the data, are plotted in Figure S10.

Table S1: Information on the field sites used to constrain spatial heterogeneity in the signal-to-noise analysis. The number of pooled cores corresponds to the number of-sub sample cores that were combined for each measured sample. Soil heterogeneity refers to the σ of log-normal fits to soil concentration distributions normalized to the field mean such that the resulting distribution has a mean of 1 (Figure S9). Site names are anonymized and location data are rounded to one decimal degree to protect farmer privacy.

								soil heterogeneity (σ; log-normal)			
Site name	**Lat**	**Lon**	**size**	**# samples**	**# pooled cores**	**sample density**	**core density**	**Ca**	**Mg**	**Na**	**Ti**
	[°]	[°]	[ha]			[ha^{-1}]	[ha^{-1}]	[]	[]	[]	[]
Site 1	45.3	-87.6	6.42	40	2	6.23	12.46	0.493	0.278	0.072	0.120
Site 2	42.3	-73.6	5.08	41	2	8.07	16.14	0.395	0.309	0.250	0.288
Site 3	31.3	-84.4	2.02	40	2	19.80	39.60	0.582	0.218	0.630	0.264
Site 4	35.8	-78.2	42.44	25	12	0.59	7.07	0.519	0.523	0.510	0.154
Site 5	35.8	-78.2	26.85	38	12	1.42	16.98	0.355	0.687	0.391	0.177

Generally, a random variable is log-normally distributed if:

$$X \sim LogNormal(\mu, \sigma) \quad (1)$$

Which means that:

$$ln(X) \sim N(\mu, \sigma^2) \quad (2)$$

where μ is the mean, σ the standard deviation, and σ^2 the variance of the respective distributions, with log-normal distributions conventionally defined via the standard deviation of the underlaying normal distribution. The expected value (mean) of a log-normal variable X can be calculated as:

$$\mathrm{E[X]} = \mathrm{e}^{\left(\mu + \frac{\sigma^2}{2}\right)} \quad (3)$$

Hence, when using the parameters of log-normal fits to populations with a given mean (Figure S9) to generate synthetic data for the Monte Carlo simulations, if generating μ and σ independently, the mean of the resulting populations will not be the same as of the initial distribution (i.e., 1). Or said differently, if we want the mean of a synthetic distribution to be a specific value, μ and σ are not independent—only one can

be randomly generated. We implement this into the Monte Carlo simulation by randomly generating shape parameters (σ_{syn}) and then calculating μ_{syn} such that $E(X) = 1$:

$$\mathrm{E[X]} = \mathrm{e}^{\left(\mu_{syn}+\frac{\sigma_{syn}^2}{2}\right)} = 1 \qquad 4$$

Now, taking the natural logarithm:

$$\ln\left(\mathrm{e}^{\left(\mu_{syn}+\frac{\sigma_{syn}^2}{2}\right)}\right) = \ln(1) \Rightarrow \mu_{syn} + \frac{\sigma_{syn}^2}{2} = 0 \Rightarrow \mu_{syn} = -\frac{\sigma_{syn}^2}{2} \qquad 5$$

The empirically constrained simulated μ_{syn} and σ_{syn} describe log-normal distributions with a mean of 1 and σ (shape) parameters constrained from field data (with a mean of 1), and are used to randomly generate sets of samples by multiplying these in-field variance factors with true "true" sample compositions.

Because the σ values from the fit to field data (Figure S9) are neither normally nor log-normally distributed (negative R^2; Figure S10), in the Monte Carlo simulations we generate synthetic σ_{syn} values by randomly pulling from uniform distributions set out by the minimum and maximum observed σ values observed in field data (for Ca, Mg, and Na the used values are 0.072402 and 0.687422, and for Ti 0.119775 and 0.288003).

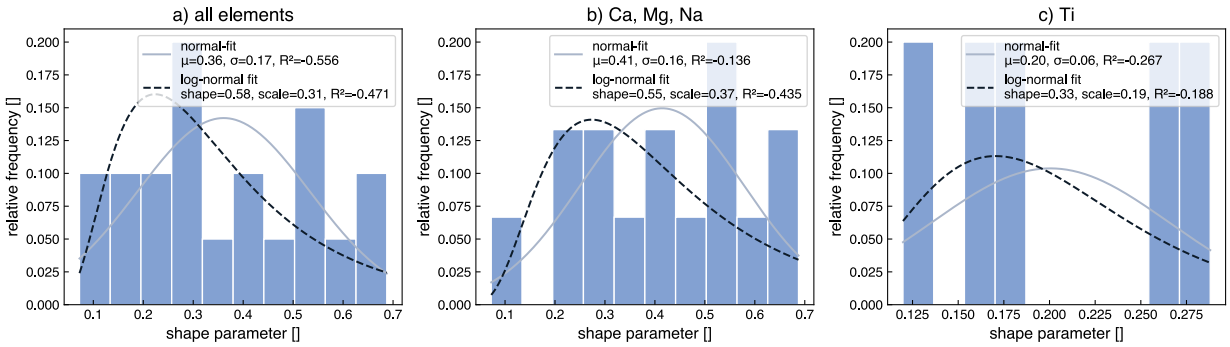

Figure S10: Histograms as well as normal and log-normal fits to the shape parameters from log-normal fits to soil data. The signal-to-noise analysis and related Monte Carlo simulations use uniform distribution set out by the minimum and maximum Ca, Mg, and Na shape values (b) as well as Ti shape values (c) due to low fit of both normal and log-normal distributions.

S2.3 Implementation of paired sampling in Monte Carlo simulations

In addition to the Monte Carlo signal-to-noise analyses that assume un-paired sampling (i.e., independently generate log-normal distributions for baseline and post–weathering soil compositions), we also simulate expected errors of the soil MRV approach when using paired samples.

Here, the work flow is adjusted for the generation of post-weathering samples (Fig S10, flowchart). As before, field data is used to generate synthetic log-normal distribution parameters for baseline samples (see methods and S2). However, for paired sampling, for each individual synthetic baseline sample the "true" post weathering composition based on the simulated feedstock application amount and dissolution fraction is calculated first. Next, we generate variance around "true" post-weathering compositions of baseline samples by generating a multiplier for each sample based on the generated log-normal shape parameter for this simulation, scaled such that the mean of the generated factors is 1 (eq. S51 (update from above)). Compared to the shape parameters used to simulate the baseline samples, the value of the shape parameter is reduced by 50% reflecting the efficacy of a paired sampling approach to reduce sampling variance. While arbitrary, this can be tested for any real deployment.

As expected, the paired sampling approach drastically reduces expected errors (Fig S11). While for individual fields, paired sampling is necessary to yield adequate errors, for aggregated monitoring lower errors are possible even when using a non-paired approach. Hence, the analysis suggests that paired sampling may not be necessary when using an aggregated monitoring approach.

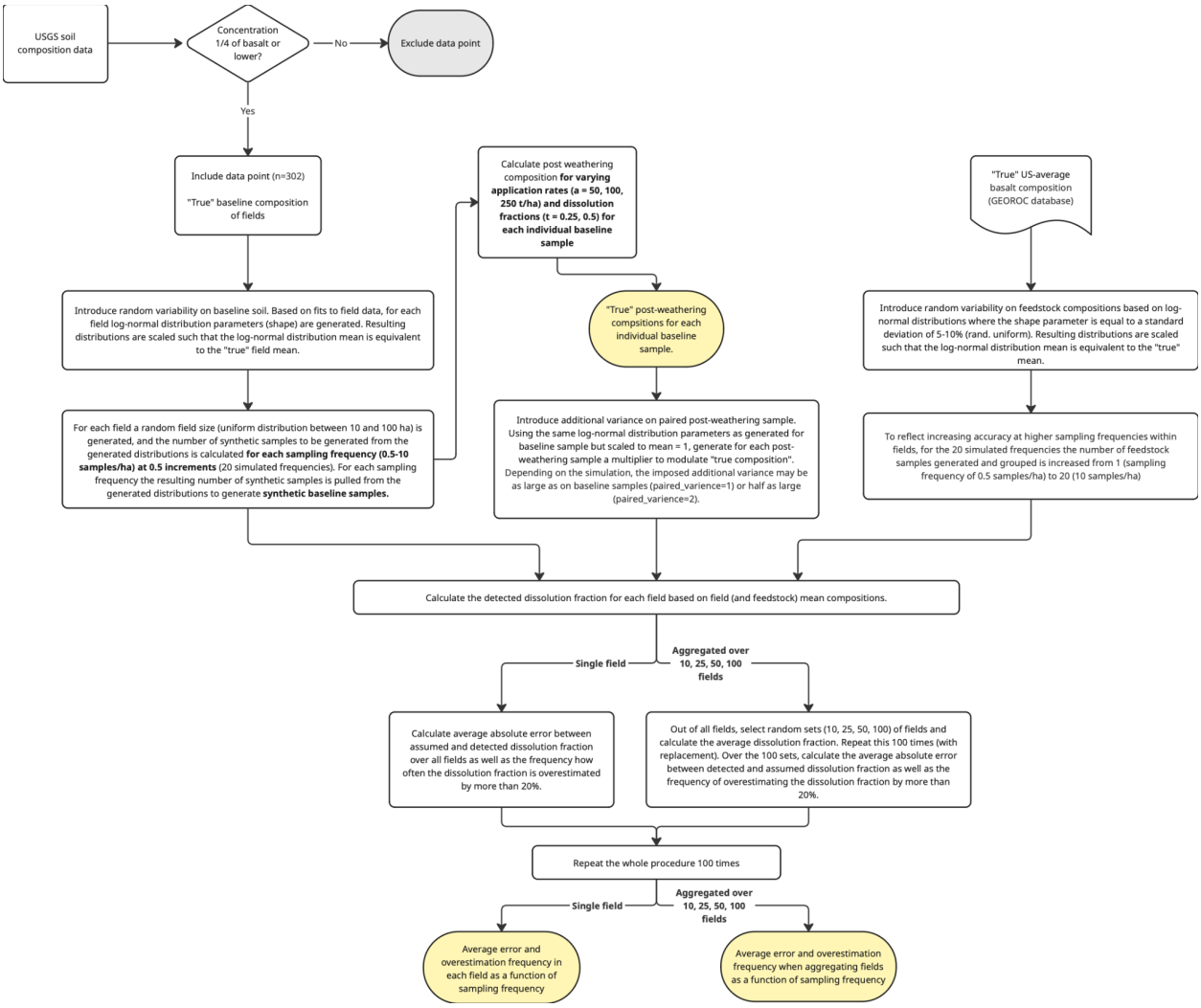

Figure S11: Flow chart for the Monte Carlo type signal-to-noise analysis using a paired sampling approach where post-weathering samples are taken at the same sites as baseline samples.

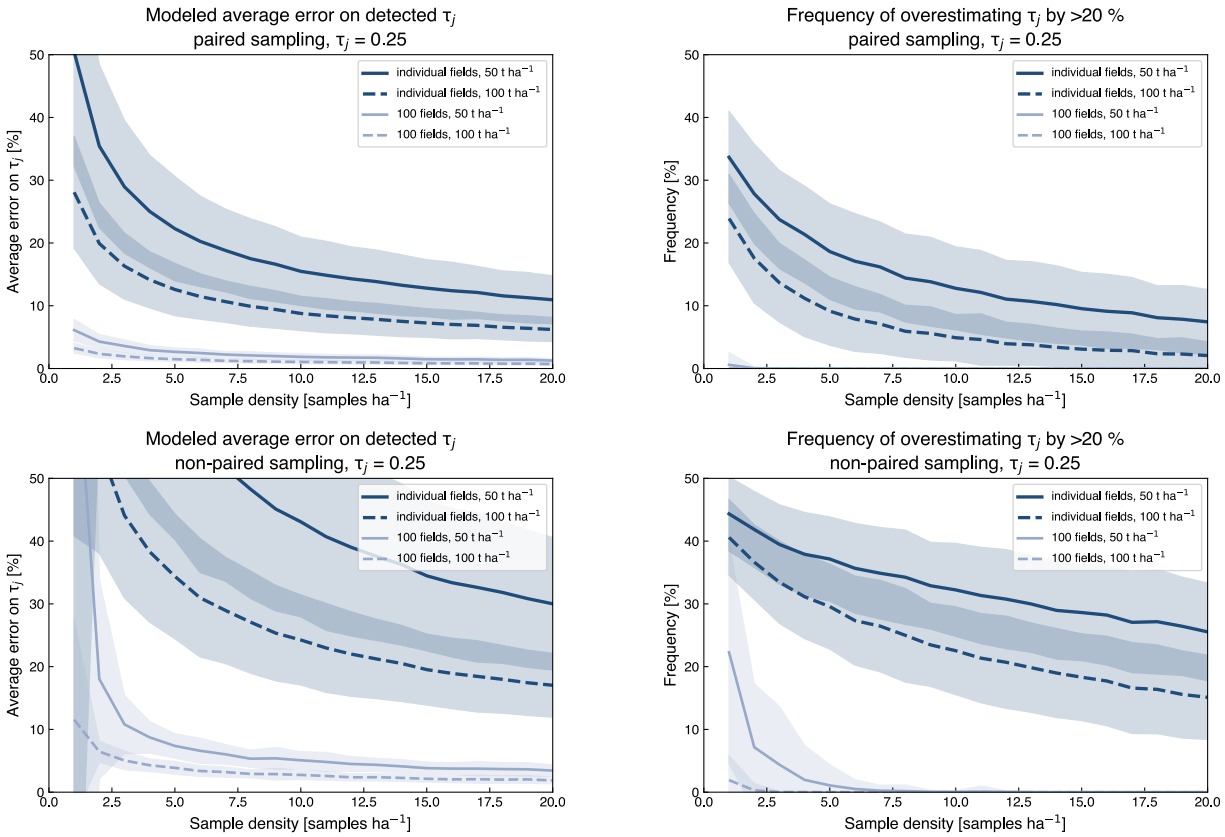

Figure S12: Comparison of selected signal-to-noise simulations between simulated to paired sampling (top row) and non-paired sampling (bottom row). Panels a and c show average error on detected mass transfer coefficients (i.e., dissolution fraction), b and d the frequency of overestimating mass transfer coefficients by more than 20% for constant τ_j but variable application amounts (a and b) as well as at constant application amount but carriable τ_j (c and d).