Geoderma xxx (xxxx) xxx-xxx

Contents lists available at ScienceDirect

Geoderma

journal homepage: http://ees.elsevier.com

Evaluation of a field kit for testing arsenic in paddy soil contaminated by irrigation water

Linden B. Huhmann [a], Charles F. Harvey [a], Jason Gross [b], Anjal Uddin [c], Imtiaz Choudhury [c], Kazi M. Ahmed [c], John M. Duxbury [d], Benjamin Bostick [e], Alexander van Geen [e,*]

[a] *Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[c] *Department of Geology, University of Dhaka, Dhaka 1000, Bangladesh*
[d] *School of Integrative Plant Science, Cornell University, Ithaca, NY 14850, USA*
[e] *Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA*

ARTICLE INFO

Handling Editor: Daniel Said-Pullicino

Keywords
Field kit
Rice paddy
Irrigation
Arsenic

ABSTRACT

Rice is the primary crop in Bangladesh and rice yield is diminished due to the buildup of arsenic (As) in soil from irrigation with high-As groundwater. Soil testing with an inexpensive kit could help farmers target high-As soil for mitigation or decide to switch to a different crop that is less sensitive to As in soil. A total of 3240 field kit measurements of As in 0.5 g of fresh soil added to 50 mL of water were compared with total soil As concentrations measured on oven-dried homogenized soil by X-ray fluorescence (XRF). For sets of 12 soil samples collected within a series of rice fields, the average of kit As measurements was a linear function of the average of XRF measurements ($r^2 = 0.69$). Taking into account that the kit overestimates water As concentrations by about a factor of two, the relationship suggests that about a quarter of the As in paddy soil is released in the kit's reaction vessel. Using the relationship and considering XRF measurements as the reference, the 12-sample average determined correctly whether soil As was above or below a 30 mg/kg threshold in 86% of cases where soil As was above the threshold and in 79% of cases where soil As was below the threshold. We also used a Bayesian approach using 12 kit measurements to estimate the probability that soil As was above a given threshold indicated by XRF measurements. The Bayesian approach is theoretically optimal but was only slightly more accurate than the linear regression. These results show that rice farmers can identify high-As portions of their fields for mitigation using a dozen field kit measurements on fresh soil and base their decisions on this information.

1. Introduction

Much of the irrigation water in rice-growing regions of Bangladesh is naturally contaminated with high concentrations of arsenic (As). When rice is irrigated with this water, As concentrations can build up in rice field soil from background levels of ~5 mg/kg to as high as 40 mg/kg (Meharg et al., 2003; van Geen et al., 2006; Saha and Ali, 2007; Hossain et al., 2008; Lu et al., 2009; Panaullah et al., 2009; Dittmar et al., 2010; Neumann et al., 2011; Javed et al., 2020). Arsenic in soil can be taken up into the rice grain, resulting in human exposure to As and associated health risks (Heikens et al., 2006; Duxbury and Panaullah, 2007; Brammer et al., 2009). In high-As regions, however, drinking water from As-contaminated wells is a much more significant exposure route (van Geen et al., 2006; Polya et al., 2008).

Soil As also decreases rice yield, and the buildup of irrigation water As in soil is estimated to reduce boro rice yield by 7–26% across Bangladesh (Abedin et al., 2002; Panaullah et al., 2009; Huhmann et al., 2017). Average boro rice yield in Bangladesh is around 4 t/ha (BBS, 2016), and with each 10 mg/kg increase in soil As, boro rice yield is expected to decrease by 0.6–1.1 t/ha (Panaullah et al., 2009; Huhmann et al., 2017). Farmers surveyed in our study area in 2016 reported receiving BDT 16 (USD 0.20) per kilogram for their rice. This implies that if rice fields were contaminated with 30 mg/kg As, for instance, and farmers could fully mitigate the negative impact of this As on rice yield, their rice yield would improve by 1.8 t/ha and their earnings would increase by about BDT 30,000 (USD 375)/ha within the mitigation area. Such elevated soil As concentrations rarely extend across an entire field, however, and are usually limited to the portion of the field closest to the inlet of irrigation water (Dittmar et al., 2010). Rice farmers considering crop-switching or other interventions

---

* Corresponding author.
E-mail address: avangeen@ldeo.columbia.edu (A. van Geen)

https://doi.org/10.1016/j.geoderma.2020.114755
Received 17 April 2020; Received in revised form 20 September 2020; Accepted 24 September 2020
Available online xxx
0016-7061/© 2020.

therefore need to know to where and to what extent their fields are affected by the buildup of As in soil. One complication in terms of obtaining a representative value is that, in spite of tilling each season, the distribution of As in paddy soil is highly heterogeneous down to very small scales. One reason is preferential accumulation in the iron plaque that coats the roots of rice plants and heterogeneous distribution of the iron plaque itself in paddy soil (Dittmar et al., 2010; Garnier et al., 2010; Seyfferth et al., 2010; Fang et al., 2018).

Various options have been considered to reduce the uptake of soil As by rice and the impacts of soil As on rice yield. These include providing cleaner irrigation water, growing As-resistant rice varieties, growing rice under conditions that are less conducive to As uptake, and removing the upper layer of As-contaminated soil (Heikens, 2006; Brammer, 2009; Polizzotto et al., 2015). Farmers have also recently begun to dig up and exchange the upper layer of paddy soil with deeper soil from the same location, placing the deeper, lower-As soil in contact with the rice roots and improving rice yield (Huhmann et al., 2019). Another alternative is to switch to other crops such as red lentil, grass pea, or coriander that are grown aerobically and may therefore not be as sensitive to the build-up of As in soil. However, farmers lack a rapid, affordable method to identify high-As soil in order to make such decisions. To provide farmers with the means to identify high-As soil with a simple method, a field kit routinely used for measuring As in water (George et al., 2012) was adapted to measure the fraction of soil As that is leachable in water. The kit measurements were validated against X-ray fluorescence (XRF) measurements of total soil As in the same soil samples. One recent study conducted a survey of As in soil from rice paddies across the Punjab plains of Pakistan, but only 103 samples were analyzed with the kit and by XRF and the statistical analysis was limited (Javed et al., 2020). Another study compared kit measurements and XRF analysis of As in 116 samples of drill cuttings from the Punjab plains of India, but these were aquifer sands very different in nature from paddy soil (Kumar et al., 2020).

2. Materials and methods

2.1. Experimental sites

The study was conducted in fields irrigated by 21 different wells in Faridpur district, Bangladesh (Fig. 1). The fields lie within the floodplain of the Padma (lower Ganges) River and are representative of silt- and clay-rich areas across the country where at least one and often two crops of rice are grown each year. The separation of 15 km at most between the fields was dictated by practical considerations, specifically the need to monitor and sample the fields throughout the growing season to study the relation between rice yield and the As content of soil (Huhman et al., 2017; 2019). In a first field that was subsequently no longer studied, the spatial variability of As concentrations in surface soil irrigated with groundwater elevated in As was documented in December 2010 by collecting 90 clumps of soil (~10 g) to 2 cm-depth on a 9 by 10 m grid and homogenizing them by kneading before analysis by XRF. This is the same field closest to the inlet of irrigation water previously studied by Panaullah et al. (2009). An additional 16 samples were collected to 2 cm-depth on a 0.25 m grid between x-coordinates 3 and 4 m and y-coordinates 1 and 2 m.

In the remaining 21 fields, irrigation wells ranged from 5 to 50 years in age and drew groundwater from 25 to 120 m depth containing 100–300 μg/L As (Table S1). Soil samples were collected and analyzed over the course of three years within 5 m × 5 m study plots whenever rice was grown. Each irrigated field contained several plots that were adjacent to each other and had been manipulated by exchanging soil or had been left undisturbed as controls (Huhmann et al., 2017; 2019). Up to two rice crops – boro (winter) and aman (summer) – were grown at most of the study sites each year. Groundwater is used for irrigation only for the boro crop in the study area; monsoonal rain is sufficient during summer. Soil was not collected during seasons when crops other than rice or no crops were grown.

2.2. Field kit soil As measurements

During each season when rice was grown, 3 soil cores 20 cm in length were collected within each 5 × 5 m study plot at roughly

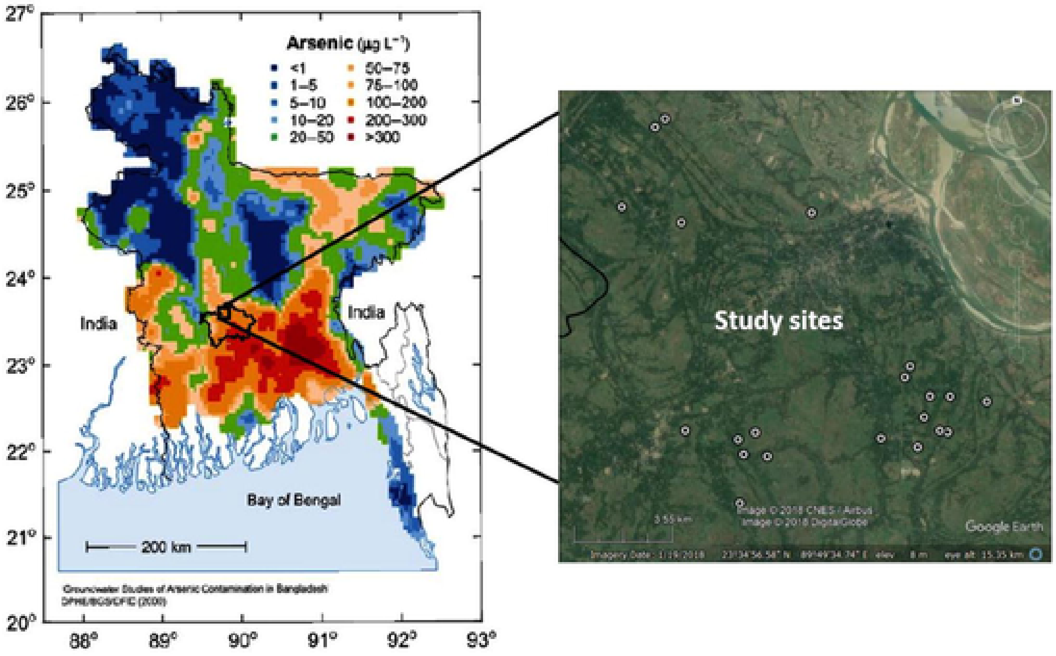

Fig. 1. Location of study sites in Faridpur, Bangladesh. Heat map of As in groundwater is from BGS and DPHE 2001.[20] Map data is from Google, CNES/Airbus, and DigitalGlobe.

monthly intervals. Each core was separated into four 5 cm subsamples. A total of 3,240 subsamples were analyzed using the ITS Econo-Quick field kit, generally in the field, but sometimes later the same day after returning to the lab.

The field kit relies on the generation of arsine gas and visual detection on a strip impregnated with mercuric bromide. The kit should therefore be used in a well-ventilated area even if the test strip is supposed to absorb the highly toxic arsine. The standard procedure for analyzing water As was adapted to measure soil As by adding 0.5 g of fresh soil, measured in the field with a portable balance, from each 5 cm interval to 50 mL of local-brand bottled water, which we have always found to be low in As. The water was confirmed not to contain any As detectable with the kit. The second reagent of the kit, an oxidant to suppress potential interference by hydrogen sulfide, was not added because it was not expected to be an issue, and the standard reaction time was maintained at 10 min. A soil kit test results in a color on the test strip that is matched to one of nine possible colored squares, each of which corresponding to a nominal As concentration referred to hereon on as a bin (Fig. S1). Extending the reaction time beyond 10 min darkens the strip further, even when water only is analyzed. No attempt was made to determine how much longer would be required to reach equilibrium because this would considerably reduce the practical usefulness of the kit. Intercalibration with laboratory measurements has shown that the kit overestimates As concentrations by about a factor of two in groundwater of the region when using the prescribed reaction time (George et al., 2012; Reddy et al., 2020).

2.3. XRF soil as measurements

The 106 soil samples collected on a grid within a single field in 2010 were analyzed in cling wrap with an Innov-X Delta Premium portable X-ray fluorescence spectrometer in the manufacturer's soil mode. The core subsamples analyzed with the kit in the field in 2015–17 were subsequently dried in an oven at 40 °C and homogenized by mortar and pestle before they were also analyzed with the same instrument for a total counting time of 35–150 s. Soil standards 2709 and 2711 from the National Institute of Standards and Technology (NIST) were analyzed at the beginning and end of each day and periodically during longer sample runs. The measured average and standard deviation for standard 2711 of 110 ± 8 mg/kg (n = 50) matched the reference value of 105 ± 8 mg/kg. The measured average and standard deviation for standard 2709 of 16.8 ± 1.6 mg/kg (n = 27) matched the reference value of 17.7 ± 0.8 mg/kg. All soil As concentrations were above the detection limit by XRF of 3 mg/kg for the shortest counting time. While high Pb concentrations can interfere with As measurements, we generally did not observe Pb concentrations higher than 50 mg/kg and obtained accurate As concentrations when making XRF measurements on NIST standard 2711 which contains more than 1000 mg/kg Pb. Since XRF measurements of total soil As concentrations have previously been validated (US EPA, 2006), these are considered the reference values for comparison with the soil kit measurements while recognizing that the kit procedure probably releases only a fraction of the total.

Kit and XRF measurements of soil As are first compared for individual samples, then averages of 3 samples per depth interval, averages of 4 samples per core, and finally plot averages of 3 cores (12 samples) per season to reduce the effect of spatial heterogeneity. Plot averages based on the kit are compared to thresholds in As concentrations measured by XRF in two different ways: on the basis of a linear conversion of all 12 plot-averaged kit measurements against 12 plot-averaged XRF measurements and by applying Bayes' theorem to different combinations of 12 individual kit readings.

2.4. Application of Bayes' theorem

Within each study plot, Bayes' theorem (Bickel and Doksum, 2001) is applied to calculate the probability that the mean XRF soil As is above a chosen intervention threshold, based on the results of m $m$kit soil As measurements. Bayes' theorem gives the probability of an event, such as plot As being above an intervention threshold, based on prior knowledge of conditions related to that event, such as specific color observations made with the soil kit. For this application, Bayes' theorem can be written as

$$P\left([\mathrm{As}]_{XRF}>\text{threshold}|bin_1,\cdots,bin_m\right)=\frac{P\left(bin_1,\cdots,bin_m|[\mathrm{As}]_{XRF}>\text{threshold}\right)}{P\left(bin_1,\cdots,bi\right.}$$

where $[\mathrm{As}]_{XRF}$ is the average soil arsenic concentration measured by XRF, bin_i bin_i is the color bin observed from a single kit measurement, and m m is the total number of kit measurements made in a study plot.

To apply Bayes' theorem, several intermediate probabilities based on the observed XRF and kit soil As measurements were calculated. First, for each kit color bin we calculate $P\left(bin_i|[\mathrm{As}]_{XRF}>\text{threshold}\right)$, the probability of getting that color if mean soil As measured by XRF is above the threshold, and $P\left(bin_i|[\mathrm{As}]_{XRF}\leq\text{threshold}\right)$, the probability of getting that color if mean soil As measured by XRF is less than or equal to the threshold (Table 1). Using this information and assuming independence of each kit measurement, the probability of getting a particular sequence of m m color bins given that mean soil As measured by XRF is above the threshold can be calculated as

$$P\left(bin_1,\cdots,bin_m|[\mathrm{As}]_{XRF}>\text{threshold}\right)=P\left(bin_1|[\mathrm{As}]_{XRF}>\text{threshold}\right)\\ \times\cdots\times P\left(bin_m|[\mathrm{As}]_{XRF}>\text{threshold}\right) \tag{2}$$

Similarly the probability of getting a particular sequence of m m bins given that mean soil As measured by XRF is below the threshold is calculated as

$$P\left(bin_1,\cdots,bin_m|[\mathrm{As}]_{XRF}\leq\text{threshold}\right)=P\left(bin_1|[\mathrm{As}]_{XRF}\leq\text{threshold}\right)\\ \times\cdots\times P\left(bin_m|[\mathrm{As}]_{XRF}\leq\text{threshold}\right)$$

The overall probability that mean As measured by XRF is above the threshold $P\left([\mathrm{As}]_{XRF}>\text{threshold}\right)$ and the overall probability that mean As measured by XRF is below the threshold $P\left([\mathrm{As}]_{XRF}\leq\text{threshold}\right)$ is calculated by counting the number of occurrences of each. The overall

Table 1
Probabilities of occurrence for each soil As bin measured by the field kit overall, and when soil As is greater than, or less than or equal to, an example 20 mg/kg threshold.

bin_i	$P\left(bin_i\right)$	$P\left(bin_i\|[\mathrm{As}]_{XRF}>20\text{mg/kg}\right)$	$P\left(bin_i\|[\mathrm{As}]_{XRF}\leq 20\text{mg/kg}\right)$
0.01	0.088	0.022	0.175
0.025	0.127	0.049	0.229
0.05	0.235	0.167	0.325
0.1	0.245	0.279	0.2
0.2	0.178	0.269	0.058
0.3	0.102	0.171	0.011
0.5	0.024	0.042	0.001
1	0.001	0.001	0.001
Sum	**1**	**1**	**1**

probability of each sequence of m bins occurring is then

$$
\begin{aligned}
&P\left(bin_1,\cdots,bin_m\right) \\
&= P\left(bin_1,\cdots,bin_m|[\mathrm{As}]_{XRF}>\text{threshold}\right) \\
&\quad \times P\left([\mathrm{As}]_{XRF}>\text{threshold}\right) \\
&\quad + P\left(bin_1,\cdots,bin_m|[\mathrm{As}]_{XRF}\right. \\
&\left.\le \text{threshold}\right) \times P\left([\mathrm{As}]_{XRF}\right. \\
&\left.\le \text{threshold}\right)
\end{aligned}
$$

All the probabilities needed to apply Bayes' Theorem are now available. These are substituted in Equation (1) to find $P\left([\mathrm{As}]_{XRF}>\text{threshold}|bin_1,\cdots,bin_m\right)$, the probability that mean As measured by XRF is above the threshold, given a sequence of m m color bins observed using the kit (Table 2).

3. Results

3.1. Spatial variability of As in soil

Concentrations of As measured on the 1-m grid (Fig. 2) ranged widely from 25 to 61 mg/kg and averaged 37 ± 7 mg/kg (n = 90, 1-sigma). For the smaller 0.25 m subgrid, the average As concentration was 42 ± 6 mg/kg (n = 16). There was no discernable trend within the gridded area akin to the pattern across a field previously documented on a larger scale (Dittmar et al., 2010; Panaullah et al., 2009).

3.2. Regression of kit As as a function of XRF As

Kit and XRF As measurements for individual 5-cm intervals were correlated during each of the five sampling seasons, with r^2 values ranging from 0.4 to 0.5 (Fig. 3). The slopes of the relation obtained by least-squares regression range for each season from 3.4 to 4.8 µg/L kit As measured in the slurry per mg/kg As in the soil measured by XRF (Fig. 3). The intercepts are not distinguishable from the origin. There is no systematic pattern in these slopes across the years or between the two seasons. The slope of the relation is slightly larger for the shallowest subsamples compared to the deeper intervals (Fig. S2). Kit measurements made on fresh soil across all seasons were combined to develop a single calibration curve (r^2 = 0.4; Fig. 4). The correlation improves when kit and XRF As values are averaged by depth interval, with 3 samples averaged per plot per season (r^2 = 0.57), or by core, with 4 samples averaged per plot per season (r^2 = 0.52). The correlation improves further when all 12 intervals collected within a plot in a season are averaged (r^2 = 0.67).

3.3. Decision rule for intervention based on linear regression

A primary goal of testing soil As with the field kit is to make a recommendation about whether farmers should intervene to mitigate the impact of soil As on their rice. One possible decision rule for intervention is to average the kit As concentrations and convert them to an average XRF concentration using the slope and intercept from a linear regression as shown in Fig. 4. If the average XRF As is estimated to be above a certain threshold, intervention is recommended.

For simplicity, we choose 20 mg/kg, 30 mg/kg, and 40 mg/kg average soil As as measured by XRF as possible thresholds for an intervention. Which threshold to use depends on the costs and the benefits to the intervening farmer, and more information about these costs and benefits would be required to choose a specific intervention threshold. Overall, among the 12-sample averages of soil As measured by XRF on our study plots, which were selected because they were impacted by As, 61% were above the 20 mg/kg threshold, 31% were above the 30 mg/kg threshold, and 19% were above the 40 mg/kg threshold.

We check how correctly the 12-sample average of kit measurements on fresh soil predicts whether a soil's mean As concentration is above or below these thresholds, using the regression equation relating the kit

Table 2
Steps in applying Bayes' theorem to calculate the probability that mean soil As is above a 20 mg/kg threshold given 12 As bins observed with the field kit.

Step Number	Description (Calculation Method)	Mathematical Expression	Example Values
(1)	Kit bin values (directly observed)	$bin_1,\cdots,bin_{12}$	0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01
(2)	Overall probability that mean As measured by XRF is above soil As intervention threshold (inferred by counting)	$P\left([\mathrm{As}]_{XRF}>20\mathrm{mg/kg}\right)$	0.567
(3)	Probability that each kit bin is observed given that mean As measured by XRF is above soil As intervention threshold (inferred by counting)	$P\left(bin_1\|[\mathrm{As}]_{XRF}>20\frac{\mathrm{mg}}{\mathrm{kg}}\right),\cdots P\left(bin_{12}\|[\mathrm{As}]_{XRF}>20\frac{\mathrm{mg}}{\mathrm{kg}}\right)$	0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022
(4)	Probability of observed kit bins given that observed mean As measured by XRF is above soil As intervention threshold (calculated by multiplying the probabilities in (3))	$P\left(bin_1,\cdots,bin_{12}\|[\mathrm{As}]_{XRF}>20\mathrm{mg/kg}\right)$	1.53×10^{-20}
(5)	Overall probability that mean As measured by XRF is equal to or below soil As intervention threshold (inferred by counting)	$P\left([\mathrm{As}]_{XRF} \le 20\mathrm{mg/kg}\right)$	0.433
(6)	Probability that each kit bin is observed given that mean As measured by XRF is equal to or below soil As intervention threshold (inferred by counting)	$P\left(bin_1\|[\mathrm{As}]_{XRF} \le 20\frac{\mathrm{mg}}{\mathrm{kg}}\right),\cdots P\left(bin_{12}\|[\mathrm{As}]_{XRF} \le 20\frac{\mathrm{mg}}{\mathrm{kg}}\right)$	0.175, 0.175, 0.175, 0.175, 0.175, 0.175, 0.175, 0.175, 0.175, 0.175, 0.175, 0.175
(7)	Probability of observed kit bins given that observed mean As measured by XRF is equal to or below soil As intervention threshold (calculated by multiplying the probabilities in (6))	$P\left(bin_1,\cdots,bin_{12}\|[\mathrm{As}]_{XRF} \le 20\mathrm{mg/kg}\right)$	8.04×10^{-10}
(8)	Overall probability of observing this set of kit bins (8) = (5)*(7) + (2)*(4)	$P\left(bin_1,\cdots,bin_{12}\right)$	3.48×10^{-10}
(9)	Probability that mean As measured by XRF is above soil As intervention threshold given that this set of bins is observed (9) = ((4) * (2))/(8)	$P\left([\mathrm{As}]_{XRF}>20\mathrm{mg/kg}\|bin_1,\cdots,bin_{12}\right)$	2.49×10^{-11}

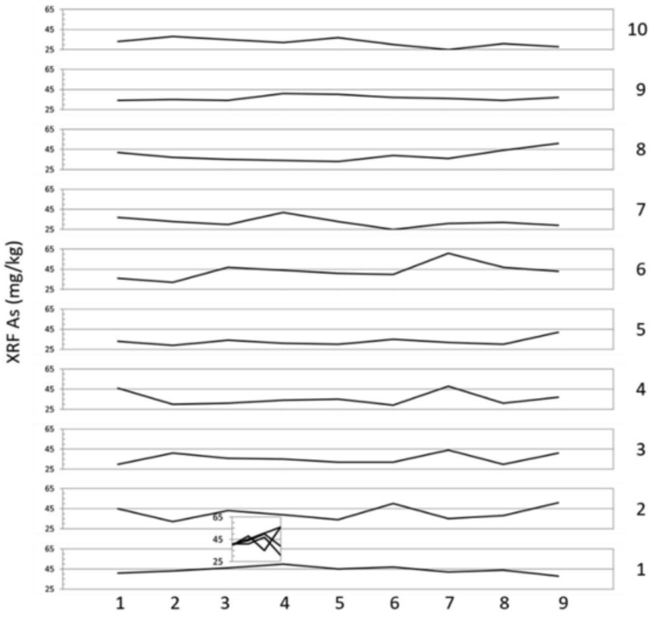

Fig. 2. Variations in surface soil As concentrations measured by XRF in Parampur in December 2010 near the inlet of irrigation water elevated in As. Concentrations of As are displayed as 10 different lines as a function of the x-coordinate for each of the y-coordinates. An additional 16 displayed samples were collected on a 0.25 m grid.

and XRF measurements and treating the mean XRF soil As concentration as the true mean soil As concentration. For a threshold of 20 mg/kg mean As measured by XRF, estimated to occur at a mean kit As of 0.09 mg/L, the kit has 85% sensitivity (true positive rate) and 81% specificity (true negative rate). That is, of the 166 study plots with more than 20 mg/kg mean soil As measured by XRF, 141 (85%) are correctly recommended for intervention by the kit, and 25 (15%) are mistakenly not recommended for intervention. Of the 104 study plots with less than 20 mg/kg mean soil As measured by XRF, 84 (81%) are correctly not recommended for intervention by the kit, and 20 (19%) are mistakenly recommended for intervention (Table 3, Fig. S3). Overall the kit measurements correctly classify 83% of the soil samples relative to the 20 mg/kg threshold.

Overall, the kit accurately classifies paddy soil with respect to intervention in 85 and 88% of cases relative to mean XRF soil As thresholds of 30 mg/kg (mean kit As of 0.13 mg/L) and 40 mg/kg (mean kit As of 0.17 mg/L), respectively. The lowest threshold (20 mg/kg) resulted in more false positives (lower specificity), where the kit incorrectly recommended intervention. The highest threshold (40 mg/kg) resulted in more false negatives (lower sensitivity), where the kit incorrectly did not recommend intervention.

3.4. Decision rule for intervention based on Bayes' theorem

The other possible decision rule uses Bayes' theorem to calculate the probability that soil As measured by XRF is above a certain threshold (e.g. 20, 30, or 40 mg/kg) based on a set of field kit measurements. In other words, a soil is judged to have an above-threshold As (for example, [As] > 20 mg/kg) if $P([As]_{XRF} > \text{threshold} \mid \text{bin}_1, \ldots, \text{bin}_{12})$ is greater than the probability threshold. If this probability threshold is too small, there is a greater possibility that the soil is erroneously judged to have an above-threshold As. If this probability is above a certain value, intervention is recommended.

For each soil As threshold, this approach allows for the selection of a probability threshold between 0 and 1, resulting in different tradeoffs between over-intervening and under-intervening. If our ideal is to intervene in plots that have an mean XRF As above the chosen soil As threshold, we can compare the number of correct and incorrect recommendations about intervention made by the kit at different probability thresholds for soil As thresholds of 20, 30, and 40 mg/kg (Fig. 5). We can then use this information about the tradeoffs to choose a preferred probability threshold for intervention.

One possibility is to select a probability threshold that $_{\text{minimi}}$zes the overall percentage of incorrect classifications, that is, a threshold that minimizes the sum of false positive and false negative kit readings. This probability threshold is 0.12 for 20 mg/kg (15.6% incorrect), 0.32 for 30 mg/kg (13.3% incorrect), and 0.82 for 40 mg/kg (11.9% incorrect).

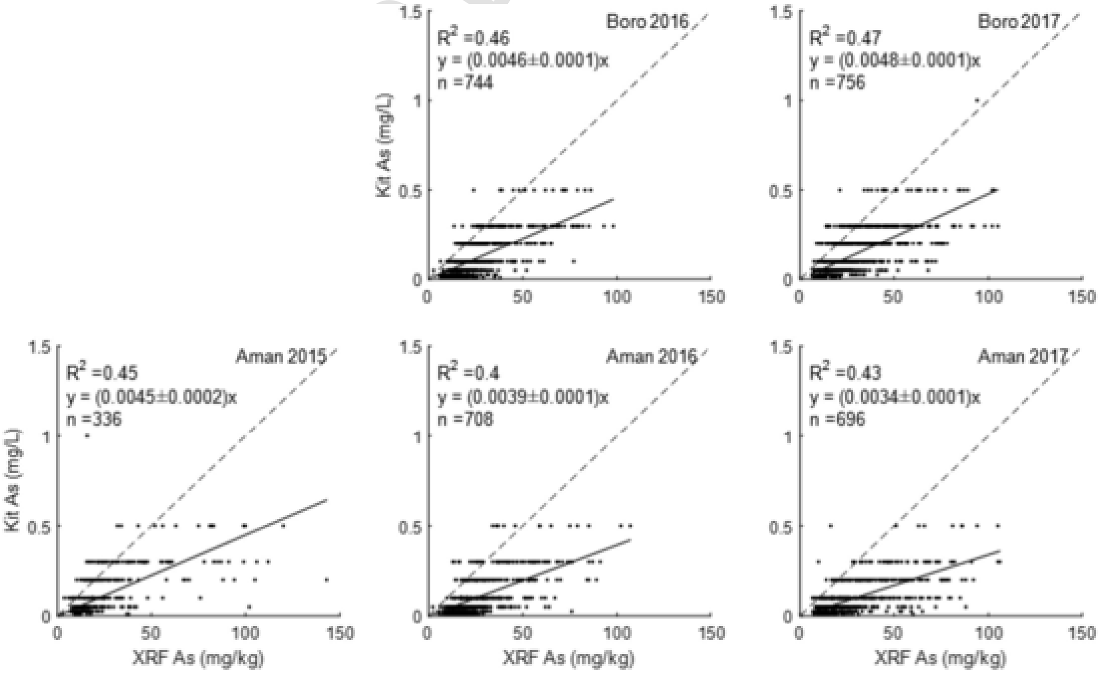

Fig. 3. Regressions of soluble As measured with the ITS Econo-Quick Field Kit as a function of soil As measured by XRF for fresh soil samples collected monthly during two boro and three aman seasons.

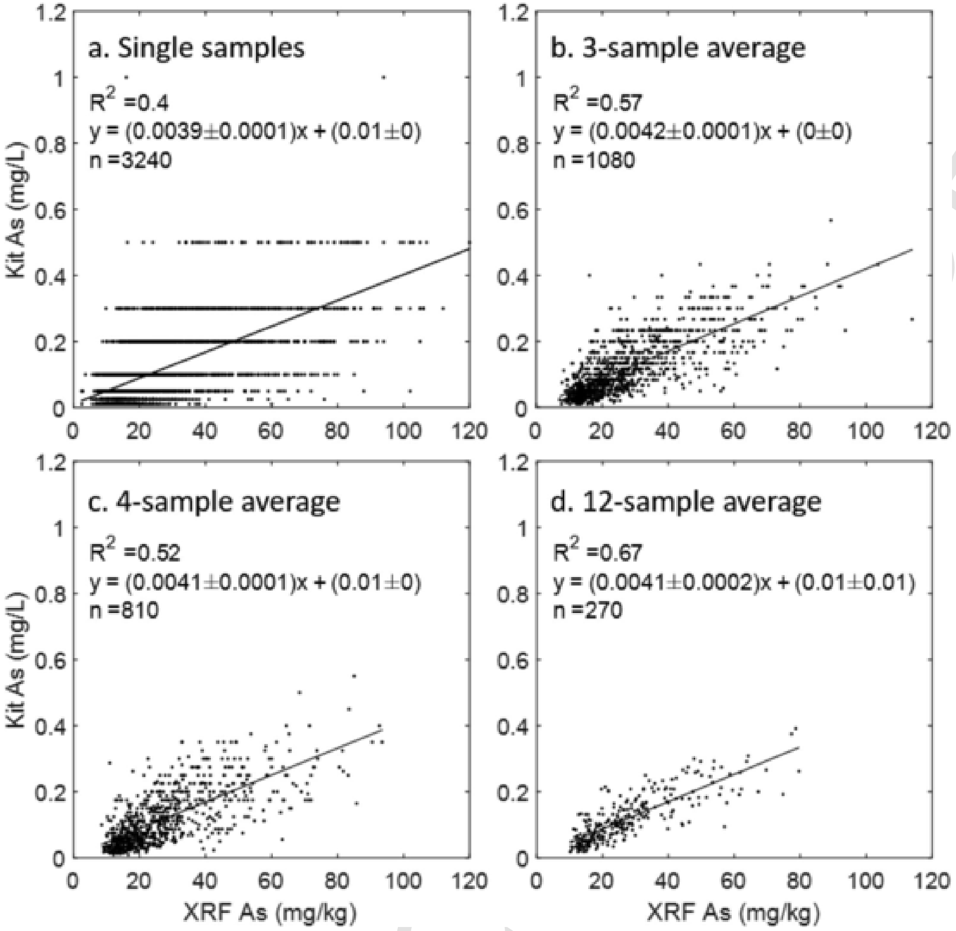

Fig. 4. Regressions of soluble As measured with the kit as a function of total soil As measured by XRF for fresh soil samples collected monthly during the first three months of two boro and three aman seasons for (a) individual subsamples at 5 cm intervals from 0 to 20 cm depth, (b) the average across the three monthly cores collected during a season for each depth interval, (c) the average of the four 5 cm subsamples for each 20 cm core, and (d) the average across subsamples and across three months for each site during each season.

Table 3
Accuracy of kit for predicting whether mean XRF soil As is above a 20 mg/kg, 30 mg/kg, and 40 mg/kg threshold for (left) the regression method (regression on an average of 12 kit samples) compared with (right) the Bayesian method (probabilities calculated from 12 kit samples) calibrated and tested on plots from the aman 2015 through aman 2017 growing seasons. For this comparison, we use the probability threshold for the Bayesian method that gives the same number of false negatives as given by the linear regression method.

		XRF				**XRF**	
		≤20 mg/kg	**>20 mg/kg**			**≤20 mg/kg**	**>20 mg/kg**
Kit (12-sample average)	**>0.09 mg/L**	20 (19%)	141 (85%)	**Kit (Probability threshold)**	**>19%**	19 (18%)	141 (85%)
	≤0.09 mg/L	84 (81%)	25 (15%)		**≤19%**	85 (82%)	25 (15%)
		≤30 mg/kg	**>30 mg/kg**			**<30 mg/kg**	**≥30 mg/kg**
	>0.13 mg/L	29 (16%)	74 (87%)		**>15%**	29 (16%)	74 (87%)
	≤0.13 mg/L	156 (84%)	11 (13%)		**≤15%**	156 (84%)	11 (13%)
		≤40 mg/kg	**>40 mg/kg**			**<40 mg/kg**	**≥40 mg/kg**
	>0.17 mg/L	23 (10%)	39 (78%)		**>82%**	22 (10%)	39 (78%)
	≤0.17 mg/L	198 (90%)	10 (22%)		**≤82%**	199 (90%)	10 (22%)

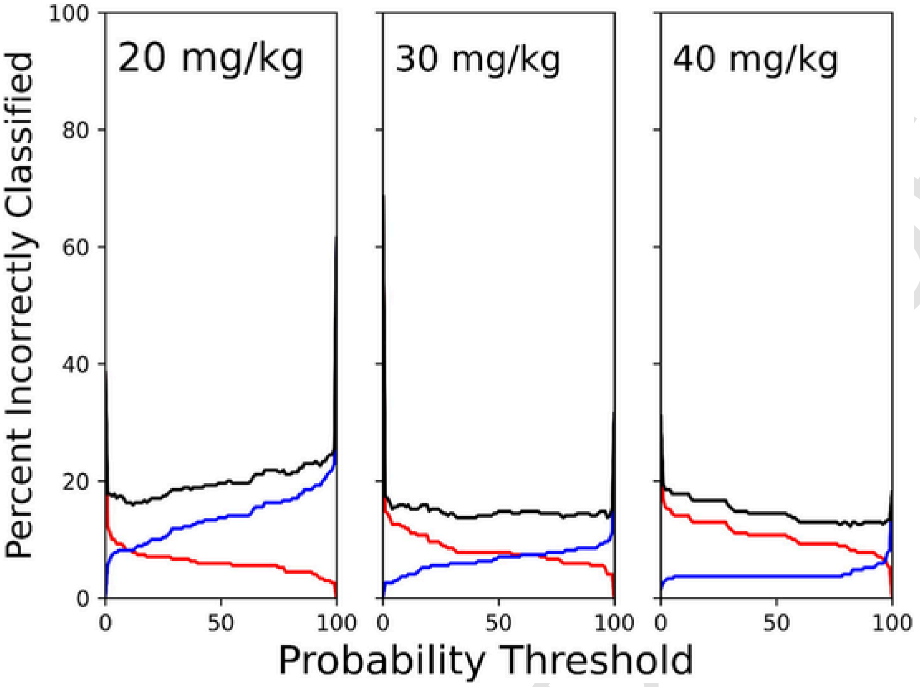

Fig. 5. Percent of samples incorrectly classified when the Bayesian method was used with the sets of 12 kit measurements to determine whether to intervene at different probability thresholds $P([As]_{XRF} > \text{soil As threshold} \mid bin_1, \ldots, bin_{12})$ for soil As thresholds of 20 mg/kg (left), 30 mg/kg (middle), and 40 mg/kg (right) . False positives are samples where the probability was above the threshold but the average XRF As was not. False negatives are samples where the probability was below the threshold but the average XRF As was not. The percent of false positives decreases when a higher (more stringent) probability threshold is used, while the percent of false negatives increases for higher probability thresholds. We also show the overall percent of samples incorrectly classified, that is, the sum of false positives and false negatives. False positives, false negatives, and overall error are all normalized by the total number of 12-sample sets of measurements.

Different value judgments could lead to other choices of probability thresholds. If farmers are willing to accept the extra effort that comes with a false positive, so long as it is very unlikely that they miss an opportunity to intervene, a lower probability threshold may be preferred. Alternatively, if farmers would prefer to sometimes miss an opportunity to benefit from an intervention, as long as they do not waste resources on a false positive, a higher probability threshold may be preferred. Thus with more information about farmers' assessments of the costs and benefits of intervening, the choice of the probability threshold for intervention could be further optimized.

4. Discussion

4.1. Spatial variability in soil as concentrations measured by XRF

Concentrations of As in the portion of the high-As field sampled at 1-m resolution often varied by a factor of 2 between adjacent samples (Fig. 2). This was the case also for the portion of the field sampled at higher spatial resolution and even for the two sides of a clump of soil that had not been homogenized by kneading (not shown). Small-scale variability was reduced in the case of XRF measurements by drying and homogenizing soil. This means a considerable portion of the scatter between kit and XRF As measurements is likely to be due to the heterogeneous distribution of As down to small scales and whether, for instance, a 0.5 g sample taken for kit analysis includes or not a portion of As-enriched Fe plaque (Garnier et al., 2010; Seyfferth et al., 2010; Fang et al., 2018). The other implication is that obtaining a representative measure of the mean As content for a portion of a field necessarily requires multiple measurements.

4.2. Variable soil As concentrations measured with the kit

An upper limit for the kit's expected response can be estimated by assuming that all As in the soil measured by XRF is released to solution. Assuming a porosity of 20% and a soil particle density of 2.5 g/cm^3, 0.5 g of fresh soil containing 50 mg/kg As could release up to 23 µg As to 50 mL of solution in the reaction vessel, thereby increasing the As concentration by 0.46 mg/L. This is a little over twice the average reading of 0.21 mg/L at 50 mg/kg As in soil (Fig. 4). The response of the kit to As in a slurry of sediment and bottled water is not known and may vary with the type of soil. Assuming the kit overestimates the As content of the slurry by a factor of two, as it does in the case of groundwater (George et al., 2012), the ratio of the predicted maximum reading to the actual reading suggests that about a quarter of the As in the soil was released by the kit procedure in this study. This is comparable to the proportion of phosphate-extractable As and phytoavailable As reported for paddy soil by Stroud et al. (2011). In two previous studies, one of wet paddy soil and one of drill cuttings analyzed with the kit, the proportion of total As in the solid phase released to solution appeared to be considerably higher, for reasons that are presently unclear (Javed et al., 2020; Kumar et al., 2020).

Over and above the variability in As concentrations measured by XRF, the regressions between soil As measured by the field kit and by XRF varied between seasons (Fig. 3). This variability between seasons could be due to factors related to the kit or the paddy soil. The test strip in the kit is sensitive to humidity, and thus results may be affected by how often and under what conditions the jar containing the test strips is left open. The characteristics of the rice paddy soil may also vary across time resulting in soil As being more or less available to the kit.

From a practical perspective, unlike the cores collected for this study to a depth of 20 cm, a farmer would typically collect multiple surface samples to establish the status of a field in relation to the loss in yield expected from high As concentrations in soil (Panaullah et al., 2009; Huhmann et al., 2017; 2019). A set of measurements obtained for one season should remain valid for several seasons because the inventory of As in the soil reflects at least a decade of accumulation. A typical farmer would not have access to a portable balance but could use the larger of the reagent spoons (volume ~1 mL, i.e. 2 g of wet soil) provided with the kit, along with a conversion to soil As adjusted accordingly. Given the high-degree of spatial variability of soil As, it is unclear that drying and homogenizing a soil sample in order to make a kit measurement on a better constrained mass of soil would provide much of a benefit. Prescribing that a soil sample must be dried before analysis with the kit, without a clear benefit, could on the other hand significantly limit adoption of the procedure.

4.3. Comparing linear regression and Bayesian decision rules

Both the linear regression and the Bayesian approach for determining when to intervene involve a calibration. In the linear regression approach, this calibration involved determining the slope and intercept of a regression between 12-sample averages of kit As and 12-sample averages of XRF As. In the Bayesian approach, this step involved calculating the probability of soil As being above an intervention threshold for

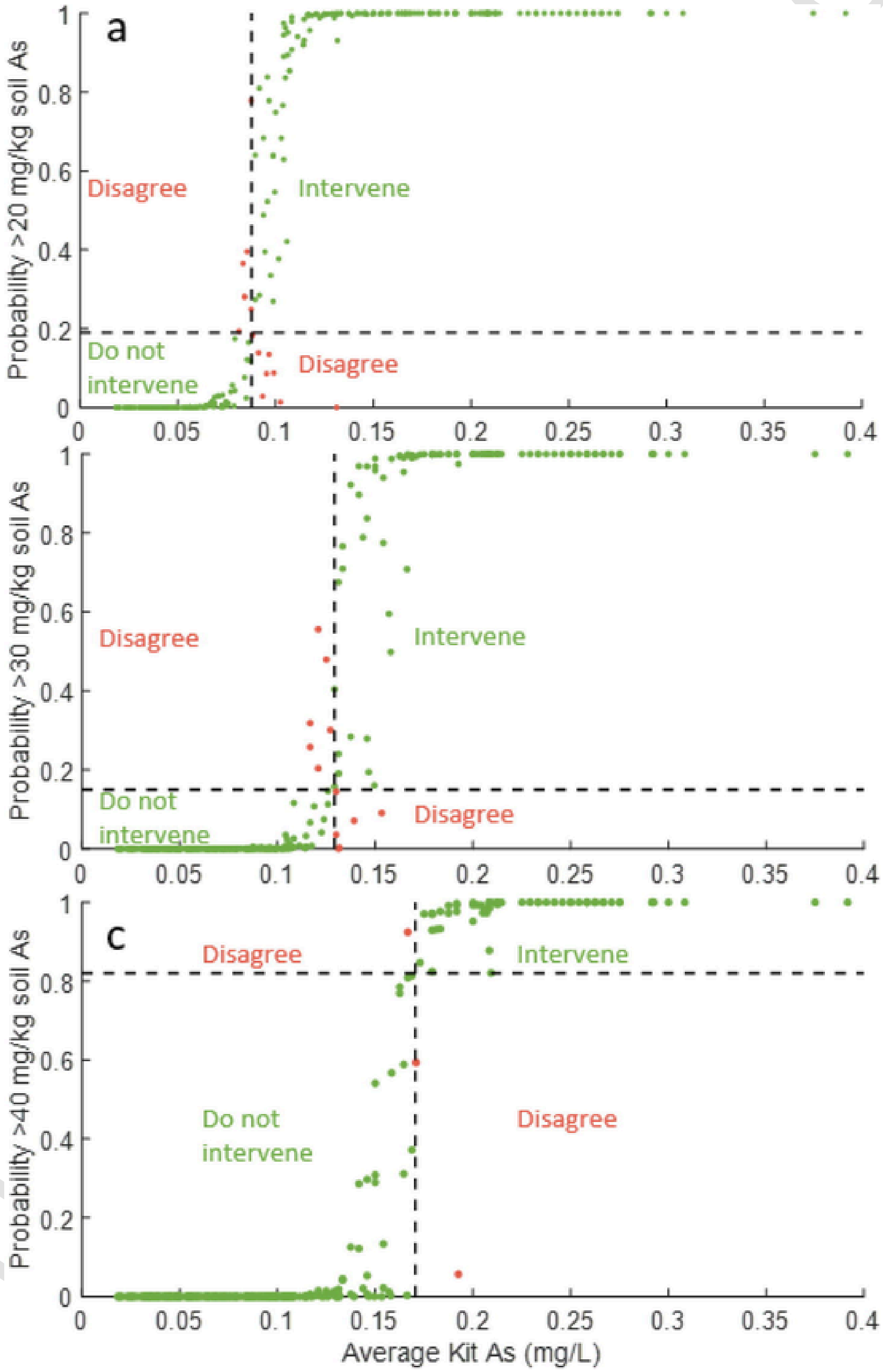

Fig. 6. Comparison of the recommendations for or against intervention made using the linear regression and the Bayesian probability approaches for intervention thresholds of a) 20 mg/kg, b) 30 mg/kg, and c) 40 mg/kg.

every possible set of 12 kit measurements. For the Bayesian approach, each soil As intervention threshold of interest requires a different table matching kit measurements to probabilities.

In the application phase, the information from the calibration phase is applied to new kit measurements collected in the field to determine whether or not the intervention threshold is reached. For the regression approach, a set of 12 kit measurements are collected and averaged, and the regression is then applied to them to estimate the average XRF As for a study plot. This average is compared to the soil As threshold to determine if intervention is recommended. For the Bayesian approach, a sequence of 12 kit measurements is matched to a pre-calculated probability of As being above or below an intervention threshold given those measurements. If this probability is above a selected threshold probability, intervention is recommended.

Comparing the calibration and application phases across the methods, we see that the Bayesian method presents two specific challenges. First, unlike the regression method, the Bayesian method requires the As intervention threshold to be known during the calibration phase. Second, during the application phase, the Bayesian method requires use of a lookup table to match a set of kit observations to a probability that soil As is above the threshold. In contrast, the regression method is just based on the average kit reading and only requires knowledge of the regression slope and intercept.

On the other hand, the Bayesian approach is the theoretically optimal approach, giving the best results (lowest error rates) possible under the set of assumptions we have made. There is no fundamental reason to believe that the field test kit values are actually linear in the true values, and using the Bayesian method does not require making this assumption. The Bayesian method utilizes all the available information and derives from scratch the relationship between kit observations and XRF measurements of As, whereas the linear regression approach starts with the values of As concentration assigned by the kit to each bin and scales them up to get estimates of XRF As. The Bayesian method is not a more reliable estimate of the concentration in soil as measured by the test kit but a more reliable estimate of what the mean XRF would be for set of soil samples. The Bayesian approach also enables a choice of a probability threshold, and thus allows for an explicit tradeoff between false positives and false negatives.

We can compare the amount of error in the recommendations from the two methods at thresholds of 20 mg/kg, 30 mg/kg, and 40 mg/kg (Table 3, Fig. S3) to see how accurate the linear regression approach is compared with the more optimal Bayesian approach. Since the errors for the Bayesian approach represent different tradeoffs at different probability thresholds, for the purposes of this comparison, we choose the probability threshold for the Bayesian method that gives the same number of false negatives as given by the linear regression method. This probability threshold is 19% at 20 mg/kg, 15% at 30 mg/kg, and 82% at 40 mg/kg soil As. Comparing the number of false positives, we find that the two methods give identical numbers of false positives at 30 mg/kg and 40 mg/kg, while the Bayesian method gives a slightly lower number of false positives (18% versus 19%) at 20 mg/kg. Overall, the linear regression and the Bayesian methods had a high level of agreement in recommending for or against intervention, with more than 90% agreement at 20 mg/kg and 30 mg/kg soil arsenic thresholds and 98% or more agreement at a 40 mg/kg soil arsenic threshold (Table S2, Fig. 6).

5. Conclusion

Analysis of a dozen samples of soil across a portion of a rice paddy with the kit can provide a farmer with a good indication of that portion's status with respect to different thresholds in average As concentration. Multiple samples would be required even for XRF measurements given the spatial variability in soil As concentrations across range of scales. The main advantage of the kit is that, outside a research setting, a portable XRF is simply not available in rural Bangladesh and won't be for the foreseeable future. The Bayesian method is more complicated to apply than the linear regression method and gives a similar or only slightly lower proportion of false positives when probability thresholds are selected that equate the number of false negatives between the two methods. On the other hand, the Bayesian method enables the selection of a probability threshold, and thus an explicit choice of tradeoffs between false positives and false negatives. If this advantage of the Bayesian method is not desired, the simpler linear regression method appears to provide sufficient guidance for or against an intervention.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by NSF Dynamics of Coupled Natural and Human Systems (CNH) grant ICER 1414131 and NIEHS Superfund Research Program grant P42 ES010349. Linden Huhmann was supported by an NSF Graduate Research Fellowship. This is contribution number 8444 of Lamont-Doherty Earth Observatory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2020.114755.

References

Abedin, M.J., et al., 2002. Arsenic accumulation and metabolism in rice (*Oryza Sativa L.*). Environ. Sci. Technol. 36, 962–968.

Bangladesh Bureau of Statistics. Estimates of Boro Rice (Husked), 2015-2016; Dhaka, 2016.

Bickel, P.J., K.A. Doksum. Mathematical Statistics: Basic Ideas and Selected Topics, Vol I, 2nd Edition. Pearson, 2001.

Brammer, H., 2009. Mitigation of arsenic contamination in irrigated paddy soils in South and South-East Asia. Environ. Int. 35, 856–863.

Brammer, H., et al., 2009. Arsenic in groundwater: a threat to sustainable agriculture in South and South-East Asia. Environ. Int. 35, 647–654.

British Geological Survey/Department of Public Health Engineering. Arsenic Contamination of Groundwater in Bangladesh Vol 1: Summary. In British Geological Survey Technical Report WC/00/19; Kinniburgh, D. G., Smedley, P. L., Eds.; British Geological Survey: Keyworth, 2001; Vol. 1.

Dittmar, J., et al., 2010. Arsenic accumulation in a paddy field in Bangladesh: Seasonal dynamics and trends over a three-year monitoring period. Environ. Sci. Technol. 44, 2925–2931.

Duxbury, J.M., Panaullah, G., Remediation of Arsenic for Agriculture Sustainability, Food Security and Health in Bangladesh; FAO Working Paper, Rome, 2007.

Fang, W., et al., 2018. Field-scale heterogeneity and geochemical regulation of arsenic, iron, lead, and sulfur bioavailability in paddy soil. Environ. Sci. Technol. 52, 12098–12107.

Garnier, J.-M., et al., 2010. Temporal variations in arsenic uptake by rice plants in Bangladesh: The role of iron plaque in paddy fields irrigated with groundwater. Sci. Total Environ. 408, 4185–4193.

George, C.M., et al., 2012. Evaluation of an arsenic test kit for rapid well screening in Bangladesh. Environ. Sci. Technol. 46, 11213–11219.

Heikens, A. Arsenic Contamination of Irrigation Water, Soil and Crops in Bangladesh: Risk Implications for Sustainable Agriculture and Food Safety in Asia. FAO RAP Publication 2006/20.

Hossain, M.B., et al., 2008. Spatial variability of arsenic concentration in soils and plants, and its relationship with iron, manganese and phosphorus. Environ. Pollut. 156, 739–744.

Huhmann, B.L., et al., 2017. Field study of rice yield diminished by soil arsenic in Bangladesh. Environ. Sci. Technol. 51, 11553–11560.

Huhmann, B., et al., 2019. Inversion of high-arsenic soil for improved rice yield in Bangladesh. Environ. Sci. Technol. 53, 3410–3418.

Javed, A., et al., 2020. Soil arsenic but not rice arsenic increasing with arsenic in irrigation water in the Punjab plains of Pakistan. Plant and Soil 450, 601–611.

Kumar, A., et al., 2020. Regulation of groundwater arsenic concentrations in the Ravi, Beas, and Sutlej floodplains of Punjab, India. Geochim. Cosmochim. Acta 276, 384–403.

Lu, Y., et al., 2009. Baseline soil variation is a major factor in arsenic accumulation in Bengal Delta paddy rice. Environ. Sci. Technol. 43, 1724–1729.

Meharg, A.A., Mazibur Rahman, M.d., 2003. Arsenic contamination of Bangladesh paddy field soils: Implications for rice contribution to arsenic consumption. Environ. Sci. Technol. 37, 229–234.

Neumann, R.B., et al., 2011. Rice field geochemistry and hydrology: an explanation for why groundwater irrigated fields in Bangladesh are net sinks of arsenic from groundwater. Environ. Sci. Technol. 45, 2072–2078.

Panaullah, G.M., et al., 2009. Arsenic toxicity to rice (*Oryza Sativa L.*) in Bangladesh. Plant Soil 317, 31–39.

Polizzotto, M.L., et al., 2015. Channels with jute-mesh structures to decrease arsenic loading to rice fields in Bangladesh. Ecol. Eng. 74, 101–106.

Polya, D.A., et al., 2008. Rice is a major exposure route for arsenic in Chakdaha Block, Nadia District, West Bengal, India: a probabilistic risk assessment. Appl. Geochem. 23, 2987–2998.

Reddy, R.R., et al., 2020. Evaluation of arsenic field test kits for drinking water: recommendations for improvement and implications for arsenic affected regions such as Bangladesh. Water Res. 2020. doi:10.1016/j.watres.2019.115325.

Saha, G.C., Ali, M.A., 2007. Dynamics of arsenic in agricultural soils irrigated with arsenic contaminated groundwater in Bangladesh. Sci. Total Environ. 379, 180–189.

Seyfferth, A.L., et al., 2010. Speciation, and co-occurrence with iron on rice (*Oryza sativa L.*) roots having variable Fe coatings. Environ. Sci. Technol. 44, 8108–8113.

Stroud, J.L., et al., 2011. Assessing the labile arsenic pool in contaminated paddy soils by isotopic dilution techniques and simple extractions. Environ. Sci. Technol. 45, 4262–4269.

US Environmental Protection Agency. Method 6200 : Field Portable X-Ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment. Test Methods Eval. Solid Waste, US Environ. Prot. Agency 2006, No. February, 1–32.

van Geen, A., et al., 2006. Impact of irrigating rice paddies with groundwater containing arsenic in Bangladesh. Sci. Total Environ. 367, 769–777.