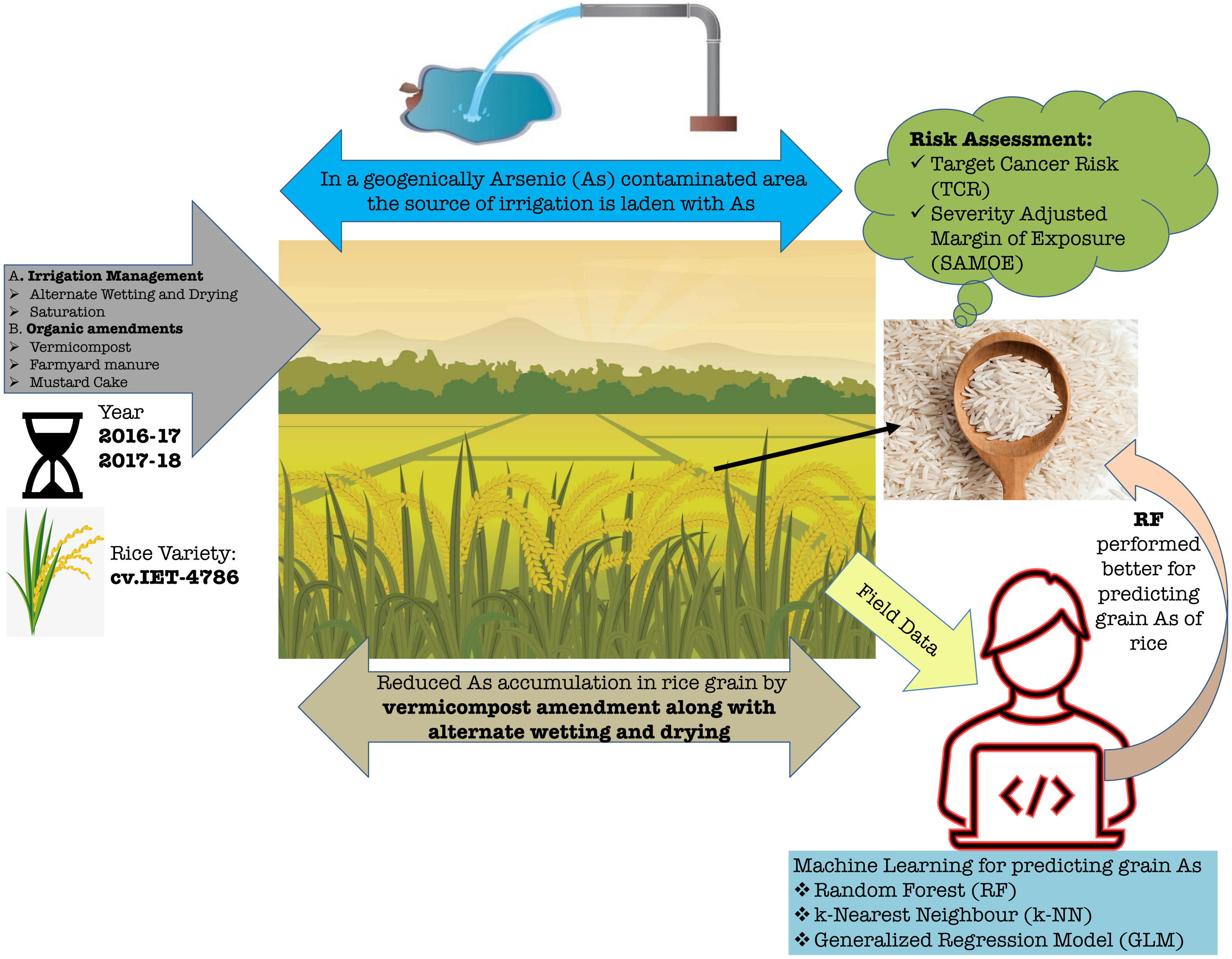
Graphical abstract



In a geogenically Arsenic (As) contaminated area the source of irrigation is laden with As

A. **Irrigation Management**
- Alternate Wetting and Drying
- Saturation

B. **Organic amendments**
- Vermicompost
- Farmyard manure
- Mustard Cake

Year
2016-17
2017-18

Rice Variety:
cv.IET-4786

Risk Assessment:
- ✓ Target Cancer Risk (TCR)
- ✓ Severity Adjusted Margin of Exposure (SAMOE)

Field Data

RF performed better for predicting grain As of rice

Reduced As accumulation in rice grain by **vermicompost amendment along with alternate wetting and drying**

Machine Learning for predicting grain As
- ❖ Random Forest (RF)
- ❖ k-Nearest Neighbour (k-NN)
- ❖ Generalized Regression Model (GLM)

Highlights

- Rice serves as the most potent dietary arsenic (As) pathway warranting mitigation.
- Alternate wetting and drying and vermicompost can reduce As in rice grain.
- Treatments were effective in reducing dietary As risk
- Random Forest can be an effective Machine Learning tool for predicting grain As.
- Paired samples, different soil and genotypes can enhance model robustness and predictability.

Deficit irrigation and organic amendments can reduce dietary arsenic risk from rice: introducing machine learning-based prediction models from field data

Sudip Sengupta[a], Kallol Bhattacharyya[a], Jajati Mandal[b,c*], Parijat Bhattacharya[a], Sanjay Halder[a] and Arnab Pari[a]

[a]*Department of Agricultural Chemistry and Soil Science, Faculty of Agriculture, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur 741252, Nadia, West Bengal, India*

[b]*Department of Soil Science and Agricultural Chemistry, Bihar Agricultural University, Sabour, Bhagalpur 813210, Bihar, India.*

[c]*School of Science, Engineering and Environment, University of Salford, Manchester, M5 4WT, United Kingdom*

[*]***Corresponding author:***jajati.bckv@gmail.com

Abstract

Dietary rice consumption can assume a significant pathway of the carcinogenic arsenic (As) in the human system. In search of a viable mitigation strategy, a field experiment was conducted with rice (cv. IET-4786) at geogenically arsenic-contaminated areas (West Bengal, India) for two consecutive years. The research aimed to explore irrigation management (saturation and alternate wetting and drying), and organic amendments (vermicompost, farmyard manure, and mustard cake) efficiencies in reducing As load in the whole soil-plant system. A thrice replicated strip plot design was employed and As content in the soil, plant parts, and the associated soil physicochemical properties were determined through a standard protocol. Results revealed that the most negligible As accumulation in the edible grains was accomplished by vermicompost amendment along with alternate wetting and drying (0.318 mg kg^{-1}) over farmer's practice of continuous submergence with no manure situation (0.895 mg kg^{-1}). Interestingly, an increase in the grain yield by 25% was also observed. The risk of dietary exposure to As through rice was assessed through target cancer risk (TCR) and severity adjusted margin of exposure (SAMOE) mediated risk thermometer. The adopted strategy made all the risk factors benign to ensure a better standard of health. The Machine Learning algorithm revealed that Random Forest performed better in predicting grain As concentration than k-Nearest Neighbour and Generalized Regression Model. Hence, if properly calibrated and validated, the former can represent an effective tool for predicting grain As concentration in rice.

Keywords: Rice grain, arsenic concentration, alternate wetting and drying, vermicompost, dietary risk assessment, Random Forest.

1. INTRODUCTION

The ubiquitous toxic metalloid arsenic (As) has sparked a number of public concerns. Its increased occurrence in the biosphere (Sanyal, 2017) is concerning from an environmental and human health perspective (Guha Mazumder et al., 2013), particularly as a persistent and group 1 human carcinogen (Menon et al., 2020). The problem of As toxicity is more severe in India and Bangladesh with groundwater As concentration several orders higher than WHO permissible limits of 0.01 mg L^{-1} (Sanyal, 2017). The drinking of contaminated water is not the sole pathway of exposure. Recent investigations have revealed that food crops, especially rice, cultivated with As contaminated irrigation water can also be a potential route of As exposure (Carrijo et al., 2019). In India and Bangladesh, daily consumption of rice is high around 68.2 and 173.3 kg $person^{-1}$ day^{-1} respectively. Approximately 69.6% of the calorific intake is from rice in Bangladesh and for India it is 29.1% (GRiSP, 2013). Rice cultivation in As-contaminated soils under anaerobic conditions results in much higher As than other crops (Awasthi et al., 2017).

The high irrigation requirement of rice contributes to the soil As build-up when irrigation water has elevated As levels (Kumarathilaka et al., 2018); thus devising a mitigation strategy should encompass both the sources. To alleviate commonly practiced flood irrigation, any drying pattern (e.g. alternate wetting and drying, AWD) can be adopted (Bakhat et al., 2017). Under AWD, flooded soils are intermittently dried to introduce periods of oxic conditions which decreases As(III) concentration in soil solution (Rahman et al., 2015). The results are highly variable unless properly adopted (Carrijo et al., 2018). Organic amendments, on the other hand, reduce As bioavailability in soils through organo-As chelation, and thus in plants, as previously stated for sesame (Sinha et al., 2011), wheat, and maize (Mandal et al., 2019b), and vegetables (Bhattacharyya et al., 2021). Since there are currently no research on the effectiveness of organic amendments in the rice environment, we decided to conduct a study that combined irrigation and organic management.

The concentration of As in rice grain should not be the only criterion for evaluating the effectiveness of interventions. The soil-crop-food transfer of As to human is vehement, so health risk assessment can be a better indicator. The risk of As to human health through food consumption can be determined by target cancer risk (TCR) and severity adjusted margin of exposure (SAMOE) (Antoine et al., 2017; Chowdhury et al., 2020).

It is necessary to determine the relationship between As in rice grain and soil properties (variables) such as pH, organic carbon (OC), available phosphorus (P), and available As. Machine learning (ML) algorithms such as k-Nearest Neighbors (KNN), Random Forest (RF) etc. can be used for this purpose. The KNN, a non-parametric classification method considers output as the average of the values of k nearest neighbors. RF as a supervised ML algorithm is widely used for classification and regression with its primary focus centered on the principle of recursive partitioning (Breiman, 2001). It is independent of the perception of functional relationships between the response and predictor variables. A comprehensive narrative of the RF algorithm can be found in Hoffman et al. (2018). RF can overcome the problem of overfitting unlike the Linear Models (LM), generalized linear model (GLM), and stepwise regression as they are less sensitive to outlier data.

With such priorities, the present study was undertaken (i) to investigate the efficacy of water management and organic amendment in lowering As levels in soil and rice edibles, (ii) to evaluate the treatments' efficacy in lowering human health risks, and (iii) to compare the efficacy of ML algorithms in predicting As in rice grain.

2. MATERIALS AND METHODS

2.1. Site features and experimental design

The field experiment was conducted in an As contaminated village, Dakshin Panchpota (23°00'N, 88°60'E) of Chakdah block of Nadia district of West Bengal, India. The site was selected based on the As the concentration of the groundwater (0.42 mg L^{-1}) used for irrigation (Referring the Village Summary of Tube-well Test Results under JPOA with UNICEF; *http://www.dngmresfoundation.org*). A typical sub-tropical climate exists in the study area with 1125-1500 mm rainfall, 40-80% relative humidity, and average maximum and minimum temperature being 37ºC and 10ºC. The investigated soil was classified as *Aeric Haplustepts*. The soil is of alluvial origin and characterized by physicochemical parameters of silty clay texture, neutral pH, and available N and K of medium/moderate concentration, high in available soil P and with high levels of As in soil and water (values of parameters are provided in subsequent section 3.1. The ratings of availability of nutrients are determined based on Supplementary Table-S1). The local popular rice (*Oryza sativa*) variety (IET-4786) was grown in experimental plots replicated thrice and laid in strip plot design with one factor as irrigation (*I_1= Saturation, I_2= Alternate wetting and drying, I_3= Continuous submergence)* and the other factor as organic amendments (*F_1=Mustard cake, F_2=Vermicompost, F_3= FYM, F_4=No manure)* in vertical and horizontal strips respectively. The experimental design is schematically represented in Supplementary Fig-S1.

2.2. Agronomic management

The experimental layout comprised of 36 plots, each 3m × 4m in size. After 3 plowings, bunds were prepared for the stagnation of water in the plots. The organic amendments (vermicompost at 3.0 t ha^{-1}, FYM at 10.0 t ha^{-1}, and mustard cake at 1.0 t ha^{-1}) were applied to 27 plots during puddling or land preparation for proper mixing with soil. In the remaining 9 plots, no organic treatments were applied. The rice seeds (cv. IET-4786) were sown in a nursery bed in the middle of December and thereafter transplanted to the main plot under the puddled condition in the last week of January with 20cm x 15cm spacing at 3-4 cm depth with 2-3 plant per hill (planting density 3,33,333 plants per hectare). In both the 2016-17 and 2017-18 study years, the same protocol was followed. The recommended dose of fertilizer of the cultivated rice variety (130:65:65 kg ha^{-1} of N: P: K) was applied. A full dose of P and K and half amount of N were applied as basal and rest N in two splits at maximum tillering and panicle initiation stage. Three levels of irrigation were applied to the respective treatment combinations as continuous submergence (*by maintaining 4 cm standing water throughout*), alternate wetting & drying (AWD) *(irrigation given on visual appearance of hair crack in experimental field)*, and saturation *(irrigation applied when soil matric potential at 15 cm depth reached -0.03 MPa after the disappearance of ponding water)*. Frequent weeding and necessary pest control measures were adopted to ensure proper growth and production of the crop. The crop was harvested in the last week of April and plant parts and root zone soils were collected from each plot leaving the edges to minimize the border effect.

2.3. Collection and preparation of soil, organics and plant samples

The initial, as well as post-harvest (PH) soil samples (0–15 cm) from the experimental sites, were collected, air-dried, ground, sieved (2-mm sieve), and finally stored in pre-marked airtight polythene packets. Standard analytical processes were adopted for physicochemical characterization. The pH of the soil was determined in 1:2 (soil: water) suspension using a combined electrode (glass and calomel electrodes) by digital pH meter (Datta et al., 1997).

Soil electrical conductivity was measured in 1:2.5 soil: water suspension (Jackson, 1973). Soil organic C was determined by Walkley and Black (1934) method; while for determining soil N, P, and K the standard methods of Subbiah and Asija (1956), Olsen and Sommers (1982), and Knudsen et al. (1982) respectively were adopted. The hydrometer method was employed for clay content determination (Bouyoucos, 1962). Soil available As concentration was determined by Olsen ($NaHCO_3$) extraction (Johnston and Barnard, 1979); while the total As was determined by following Sparks et al. (2006). The organic treatments used in the study were analyzed for their C, N, P, K, and As concentration based on the standard protocol following Page et al. (1982).

The plant (rice) samples were collected at harvest, washed initially by tap water followed by dilute hydrochloric acid, and finally with double-distilled water. The samples were then appropriately labeled, chopped, separated into the root, shoot, and grain, and dried in an air-oven at 105°C for 24 hours. The dried samples after cooling were ground and digested with a mixture of acids *i.e.* HNO_3, $HClO_4$, and H_2SO_4 in a proportion of 10:4:1 (v/v) (Jackson, 1973) and filtered using Whatman No. 42 filter paper.

2.4. Instrumental analysis

Standard analytical procedures were adopted for the determination of As in plant digest and soil extract by sequentially diluting with distilled water, reacting with concentrated HCl, KI, and ascorbic acid for 45 minutes, and then analyzing through Atomic Absorption Spectrophotometer (AAS) (Sparks et al., 2006).

Validation of the analytical methodology of As determination was made through the National Institute of Standards and Technology (NIST) prepared standard reference material of rice (SRM1568a). In comparison to the certified value of 290 ± 30 µg kg^{-1} for SRM1568a, the current Perkin Elmer AAnalyst 200 AAS attached with Flow Injection for Atomic Spectroscopy (FIAS) Systems at λ_{max}=193.7 nm exhibited As concentration as 287 ± 8.1 µg

kg^{-1}, thereby showing good agreement. Accuracy validation was done in triplicates and in every batch of 30 samples, two blank reagents and one standard reference material were analyzed.

2.5. Risk assessment of dietary exposure to As through rice grain

2.5.1. Target Cancer Risk (TCR)

TCR assumes great significance in dietary risk assessment as it categorizes the lifetime exposure of carcinogenic As for human individual. The TCR calculation is based on the following equation (Antoine et al., 2017; Bhattacharyya et al., 2021):

$$\mathbf{TCR} = \frac{\mathrm{Efr} \times \mathrm{Ed} \times \mathrm{Fir} \times \mathrm{C} \times \mathrm{CPSo}}{\mathrm{BWa} \times \mathrm{ATc}} \times 10^{-3}$$

where,

Efr = the exposure frequency to As (365 days),

Ed = the exposure duration (70 yrs)

FIR = the food ingestion rate in grams per day

C = the inorganic As concentration

CPSo = the oral cancer slope for arsenic as 1.5 (mg kg^{-1}) day^{-1}

BWa= the body weight of 68 kg

ATc= the averaged carcinogenic exposure time (365days*70yrs)

10^{-3}= the unit conversion factor (Antoine et al., 2017).

The acceptable range of TCR varies from 10^{-4} to 10^{-6}, (*i.e.* 1 in 10,000 to 1 in 1,000,000) (Shaheen et al., 2016).

2.5.2. Risk thermometer and SAMOE (Severity Adjusted Margin of Exposure)

According to the Swedish National Food Agency, a risk thermometer is an established holistic and new protocol on risk characterization (Sand et al., 2015). The risk thermometer mainly estimates the exposure of As in food and compares the health-based Tolerable Daily

Intake (TDI). The human dietary exposure of As through rice consumption is calculated using the following equation (Chowdhury et al., 2020):

$$\mathbf{SAMOE = TDI / (AF_{BMR} \times AF \times SF \times E)}$$

where,

TDI = 3.0 μg kg^{-1} $bodyweight^{-1}day^{-1}$ value for As

AF_{BMR} = Non-linear relation in dose range (1/10; BMR - Benchmark response)

AF (Assessment factors) = a factor of 10 (conservative assessment)

SF (Severity factor) = 100 (For cancer, the most severe category)

E= Different exposure factor (here, inorganic As concentration).

Based on the SAMOE value, the classes of risk in risk thermometer are prescribed, as, class 1 (no risk, >10); class 2 (no to low risk, 1-10); class 3 (low risk, 0.1-1); class 4 (moderate to high risk, 0.01-0.1) and class 5 (high risk, <0.01) (Sand et al., 2015).

2.6. Statistical Analysis and Machine Learning

The data collected for two years on soil and grain chemical properties were initially subjected to Shapiro-Wilk normality test. On confirmation of normalization, the mean effects were compared with Duncan's multiple range test. Apart from these, simple descriptive statistics (mean, standard deviation, etc.), prediction modeling, risk assessment of As through rice were performed using *Microsoft Excel 2016* and R-Studio (*Version 1.3.1093 2.3.1*).

2.6.1. Random Forest

Random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution using voting. It is an ensemble method that is better than a single decision tree because it reduces the over-fitting by averaging the result. The variable importance function within the RF algorithm ranks predictor variables based on the increase in model error by randomly permuting the values of the predictor variables. Briefly, the mean square error (MSE) for each tree is the average squared deviations of MSE observations from the predictions. Here we have used the package *Random Forest (version 4.6-14*) for analysis. The whole data set (n=36) was used for the purpose with 10-fold cross-validation repeated 5 times by using the *package caret (version 6.0-86).* The *mtry=4* and *ntree=1000* resulted in the minimum Root mean Squared Error (RMSE) and maximum R^2 value and was selected as the final model.

2.6.2. k-Nearest Neighbors

In non-parametric KNN regression, the output is the property value for the object (Evelyn and Hodges, 1951; Altman, 1991). This value is the average of the values of k nearest neighbors. Given a value for *k* and a prediction point x_0, KNN regression first identifies the *k* training observations that are closest to x_0, represented by N_0. It then estimates f(x_0) using the average of all the training responses in N_0. Mathematically it can be represented as follows (Song et al., 2017):

$$\mathrm{f}(x0) = \frac{1}{k}\sum_{x€N0} yi$$

The whole data set (n=36) was used for the purpose with 10-fold cross-validation repeated 5 times by using the *package caret (version 6.0-86).* The *k=3* resulted in minimum Root mean Squared Error (RMSE) and maximum R^2 value and was selected as the final model.

2.6.3. Generalized Linear Models

Generalized linear models (GLM) allow the extension of linear modeling ideas to a wider class of response types, such as count data or binary responses. GLM fits models of the form g(Y) = XB + e, where the function g (Y) and the sampling distribution of e need to be specified. The GLM unifies various other statistical models, including linear regression, logistic regression, and Poisson regression (Nelder and Wedderburn, 1972). The whole data set (n=36) was used for the purpose with 10-fold cross-validation repeated 5 times by using the *package caret (version 6.0-86).* The minimum Root means Squared Error (RMSE) and maximum R^2 value were selected as the final model.

2.6.4 Model Performance metrics

In this study, coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were calculated to assess the performance of the models. The objective is to develop a model with high performance and less error.

$$R^2 = 1 - \frac{\sum(a_i - b_i)^2}{\sum(a_i - \mu_a)^2} \tag{1}$$

Where a denotes the output values, b denotes the real values, and μ_a is the mean value of the a values, i^{th} is the number of observations such as *1,2,3…, n*.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - b_i)^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|a_i - b_i| \tag{3}$$

3. RESULTS AND DISCUSSION

3.1. Characteristics of experimental site and organic amendments

The experimental soil had a neutral pH (6.96), a low soluble salt concentration (EC- 0.42 dS m^{-1}), medium organic carbon content (0.55 %), 49 percent clay, moderate in available nitrogen (260 kg ha^{-1}) and available potassium (227 kg ha^{-1}) content, and a high available phosphorus content (32.9 kg ha^{-1}). The values of total and Olsen extractable As were relatively higher corresponding to 28.78 and 4.07 kg ha^{-1}, respectively. Normally rice is grown here by irrigating using shallow tube-well drafted As contaminated underground water (0.42±0.09 mg L^{-1}) and As concentrations in rice grain ranges from 0.785±0.164 mg kg^{-1}. Further the organic treatments were characterized for their C-N-P-K concentrations (on a percent dry weight basis), with the results of FYM, vermicompost and mustard cake used in the experiment were 14.75-0.56-0.30-0.46, 22.5-1.2-0.21-0.59 and 25.4-3.9-1.93-1.67 respectively. The C:N ratios of the treatments were 26.3:1, 18.7:1 and 6.5:1 respectively. The organic treatments used in the present study were previously tested for their As concentration before field application. The As concentration in all the organics was found to be below the detectable limit. The study area is a previously reported As contaminated site (Mukhopadhyay and Sanyal, 2004) and As uptake via rice having associated dietary risk has been confirmed by Sinha and Bhattacharyya (2014) and Chowdhury et al. (2018).

3.2. Arsenic accumulation in rice

Arsenic accumulation in rice (IET-4786) reduced significantly with a lesser extent of irrigation. It was observed that ensuring a longer dry spell by keeping the field saturated and using an alternative wetting and drying moisture regime resulted in lower accumulation of As. Arsenic in rice grain of 0.741 mg kg^{-1} (saturation) and 0.655 mg kg^{-1} (alternate waiting and drying) were recorded as opposed to the higher concentration of 0.885 mg kg^{-1} (continuous submergence) for pooled data of two years (Table 1). Under various realms of deficit irrigation, the concentration of As in the plant's root and shoot was also found to be lower (Table 1). Percent reduction in As accumulation in rice grain is depicted in Fig. 1.

According to previous reports, when rice is grown in anoxic (flooded) conditions, it takes up 10-15 times more As than when it is grown in oxygen-rich conditions (Hua et al., 2011). Reduced As in rice grain can be achieved as a result of the periodic oxidized condition caused by AWD (Li et al., 2019). Deficit irrigation mediated As reduction has been established in earlier studies. According to Mukherjee et al. (2017), the As content of polished rice decreased by 17.6–25% due to deficit irrigation. Fernández-Baca et al (2021) recorded a similar reduction of 25% As in rice grain under AWD. Shrivastava et al. (2020) put forward the reason for the low entry of As in plant due to the lower frequency of plant exposure to As by irrigation water under AWD.

The use of organic amendments was also found to significantly reduce As accumulation in rice grain as well as in root and shoot over no manure condition. The efficiency of the organic amendments to reduce As load followed the trend of vermicompost > mustard cake > FYM with values of 0.534, 0.669, and 0.768 mg kg^{-1} As, as against 0.844 mg kg^{-1} for the no-manure condition, as evident in Table-1 and through percentage in Fig 1. Organic amendment stability over the two-year period was critical. To address this issue similar treatments (organic amendments) were applied on the same plots for both years at a uniform rate. The comparison between the two years was found to be statistically non-significant in a paired T-test (Fig. 2).

Efficacy of various types of organic amendments in reducing As load in plant edibles was documented in a variety of crops in previous studies. Mandal et al. (2019b) found similar cases of reduction in As in wheat and maize grain with sugarcane bagasse > paddy husk > rice straw > vermicompost > FYM. Organic treatments were also effective in lowering the amount of As in cauliflower, spinach and tomato in the trend of vermicompost > mustard cake > FYM (Bhattacharyya et al., 2021). In all cases, a possible organo- As complexation in soil was propounded as the underlying reason for reduced uptake in plant edibles.

The interaction of the different moisture regimes and the organic amendments brought about significant changes in As accumulation in crop edibles over their controlled counterparts. The least As accumulation in the edible grains was facilitated by vermicompost amendment along with alternate wetting and drying (0.318 mg kg^{-1}) over farmer's practice (0.895 mg kg^{-1}). Such effect was also evident from the studies of Rahaman et al. (2011) and maybe adopted as a successful mitigation strategy. The current study primarily focused on the effect of the treatments on the As in grain, while the year effects on the rice grain As was found to be statistically non-significant (Fig. 2).

3.3. Dietary risk assessment to As contaminated rice grain intake

The toxicity reported in the non-endemic areas has been a growing threat and the critical evaluation of its pathway through the food chain has necessitated a thorough assessment of exposure to dietary risk. Rice is the principal dietary component in the study area and usually consumed three times a day along with vegetables (Signes-Pastor et al., 2008). The presence of As in food especially in rice samples from the West Bengal region and its health effects have already been envisaged, especially when cooked using contaminated water (Upadhyay et al., 2019). The dietary As risk is more emphasized on inorganic As (iAs) concentration which depends on variety and location. The genetic basis of As tolerance and accumulation in the early seedling stage of rice are the primary reasons behind the varietal differences as evident from the quantitative trait locus (QTL) mapping study (Murugaiyan et al., 2019). The current variety IET-4786 was reported to contain 86.6% iAs (out of total As) in the study area by the same research group (Sinha and Bhattacharyya, 2020), and the same data was used here to derive holistic expressions of human dietary risk through consumption of the contaminated rice grain. The assessment was primarily carried out in brown rice thus all risk parameters were derived based on the Joint FAO/WHO Expert Committee on Food Additives recommended maximum level of 0.4 mg kg^{-1}.

3.3.1. Target Cancer Risk (TCR) of As through rice grain

The results reported in Table-2 and details in Supplementary Table S2 about TCR through consumption of contaminated rice grain suggest that in all cases the risk associated with cancer is high. The traditionally followed farmer’s practice, as mentioned earlier, has risk (TCR- 6.64×10^{-3}), much higher than the tolerable limit of 10^{-4} (Shaheen et al., 2016). Even after the adoption of all sorts of irrigation and organic interventions through the present study, the lowest value (TCR- 2.36×10^{-3}), was, still higher than the tolerable limit. Still, AWD adoption along with vermicompost application could significantly curtail the load of As and thus ensure some safeguard against the carcinogen. The higher risks of cancer occurrence through dietary As exposure have earlier been reported (Mondal et al., 2010; Halder et al., 2014).

3.3.2. Risk thermometer and SAMOE (Severity Adjusted Margin of Exposure)

The ‘Risk thermometer’ and the calculated ‘SAMOE’ value for As toxicity through cultivated rice under water and organic management protocols showed varying concern levels of risk from class 4 (moderate-high) to class 3 (low risk) depending on As concentration (Table 2 and Fig. 3; and in details in Supplementary Table-S3). The farmer’s practice of continuous submergence without manure showed the highest SAMOE (0.04) while managing AWD with vermicompost had the least (0.112). The varying levels and origin of As in the rice grains under different interventions can cumulatively aggravate the toxic load when they enter the dietary pathways by cooked, parboiled, or even raw grain (Chowdhury et al., 2020). Consumption of contaminated rice grain as a major staple diet in conjunction with other dietary ingredients on a prolonged basis leads to As poisoning in the contaminated belts of West Bengal (Chowdhury et al., 2018; Biswas et al., 2019).

3.4. Effect of irrigation management and organic amendments on yield of rice grain

The adoption of any intervention to curtail the As load faces the major hurdle in terms of how far it is adaptable in the farmer’s field in terms of yield or the monetary return. In the current study, the applied irrigation management and organic amendments significantly increased the grain yield of the variety IET-4786, although the variations were less conspicuous than the As load curtailment. In comparison to the conventional practice of continuous submergence without manure (2.589 t ha^{-1}), AWD in conjunction with vermicompost application resulted in a higher grain yield (3.506 t ha^{-1}). Such an increase in grain yield was also observed in the single effects of water management and organic interventions (Table 1).

AWD can augment crop yield (Carrijo et al., 2018) when deeper soil layers (25–35 cm depth) have sufficient water supplying capacity during the drying periods to meet transpiration demands. The use of organics sustains soil fertility through the release of nutrients (Reddy and Reddy, 1999) and thus favor crop yield.

3.5. Effect of irrigation management and organic amendments on post-harvest soil

The effect of organic amendments and irrigation management on the post-harvest soil parameters can be observed in Table 3. The pH of the soil ranged from 6.81 for treatment I_2F_2 to 7.01 for treatment I_3F_4 and was found to be statistically different. The decrease in pH after submergence is probably due to the accumulation of CO, produced by the respiration of aerobic bacteria because CO depresses the pH even of acid soils (Kumari et al., 2021). The OC concentration of the post-harvest soil varied significantly across the treatments. The treatment I_2F_2 having OC of 0.59% recorded the highest and I_3F_4 (0.47%) recorded the lowest. Regardless of the irrigation management techniques used, the highest OC percent was found in vermicompost, followed by FYM, and Mustard Cake. The available soil P also varied significantly across the treatments. The maximum, 36.2 kg ha^{-1} was recorded in I_2F_2, and the lowest, 30.1 kg ha^{-1} was recorded in I_3F_4. The increase in available P was observed in all the treatments as the organic amendments served as a potent source of P. The nutrient status of the organic amendment as described earlier justified the above fact.

A significant reduction in the soil As was observed across the treatments. The efficacy of the organic amendments and irrigation management techniques in reducing the available As in soil followed the order $I_2F_2>I_1F_2>I_2F_1=I_3F_2>I_1F_1=I_2F_3>I_2F_4= I_1F_3>I_1F_4>I_3F_1> I_3F_3>I_3F_4$. A large number of studies established that the application of organic manure immobilizes, adsorbs, binds, or co-precipitates As *in-situ* which in turn can influence the presence, availability, and mobility in soils and aquatic environments. The complexation of Arsenite (As^{3+}) with the Humic Acid (HA) through phenolic, carboxylic, amino, and sulfhydryl functional groups, may serve as the binding sites for As by forming negatively charged adducts (Mandal et al., 2019a; Kumar et al., 2021). The direct association of As with these functional moieties may exhibit varying strength and thus represent different binding mechanisms. Furthermore, these functional groups in HA may bind As via a cation (e.g., Fe) bridge binding mechanism by forming Dissolved Organic Matter (DOM)-cation–As complexes (Ritter et al., 2006). In 2012, Ghosh et al. observed that HA/FA extracted from compost was found to be better in scavenging arsenate, and Sinha and Bhattacharyya (2011) observed higher stability of As-HA/FA complexes with vermicompost rather than FYM or oil cakes along with the formation of complexes with particulate organic matter.

3.6. Performance of the Machine Learning based Models

A comparison of the performance matrices of the models was depicted in Table 4. The results showed that RF (0.065) had the lowest RMSE, followed by KNN (0.066), and GLM (0.086). The MAE also followed the same trend as RMSE. The MAE of RF (0.055) was minimum followed by KNN (0.056) and GLM (0.070). In terms of R^2, the models followed the order KNN (0.88) > RF (0.86) > GLM (0.77). It was observed that the RF performed better in terms of RMSE and MAE compared to KNN and GLM although the R^2 of KNN was greater than RF. The RMSE measures indicate the absolute fit of the model to the data, that is how close the model's predicted values are to the actual or observed data points. While R^2 is a relative measure of fit, RMSE provides an absolute measure. As the square root of the variance, RMSE can be interpreted as the standard deviation of the unexplained variance and has the useful property of being in the same units as the response variable. Lower values of RMSE indicate a better fit. RMSE is thus the best measure of the prediction model. The significance of RMSE exists in a way that even a model with low R^2 can be practically useful if the RMSE is low (Tropsha, 2010; Alexander et al., 2015) thereby establishing the importance and significance of cross-validation (Saha et al., 2021). The same explanation goes for the MAE. The less the MAE or MAPE, the better will be the prediction by a model. MAE is the mean or average of all absolute errors between the observed and the predicted values. Hence in terms of predictability, the RF can be used for predicting the As concentration in rice grain over KNN and GLM models. The variable importance plot from

RF as can be observed from Figure 4 revealed that among the soil parameters, the soil As has the highest importance followed by pH, OC, and soil P concentration. The importance parameter was calculated in terms percentage increase in Mean Squared Error of predicting the dependent variable. This shows how much our model accuracy decreases if we leave out that variable. The top variables contribute more to the model than the bottom ones and also have high predictive power (Kuhn, 2008). The significant effect of soil As on rice grain As as observed was because As from soil was translocated to root and finally grains. So the order of As concentration in the rice plant parts were as follows root > shoot > grain (Table 1). Several authors have reported the fact that irrigation water significantly contributes towards the build-up of As in soil and in turn increases the bioavailability (Golui et al., 2017; Mukherjee et al., 2017).

4. CONCLUSION

The widespread use of As contaminated water for irrigating the crops results in the substantial entry of the contaminant in the human food chain and leads to severe health hazards. Rice, being the predominant dietary component, it's cooking with contaminated water further escalates the problem. The dietary risk parameters, that have been calculated here, envision the aggravating health hazards associated with its consumption. The traditional agricultural practice of continuous submergence and no manure application resulted in substantial entry of the carcinogen in rice grain and human diet. On the contrary, mitigation techniques in the form of irrigation management and organic amendments reduced accumulation of As in crop edibles and post-harvest soils and precisely in some of the treatment combinations made risk parameters (TCR, SAMOE) somewhat benign. Adoption of AWD and vermicompost application appeared most effective. The use of ML algorithms revealed the fact that in terms of model performance matrices RF > KNN > GLM. So, Random Forest (RF) algorithm can be used for the prediction of grain As concentration. The soil As was observed as the most important variable affecting the grain As concentration. This study will serve as proof of the efficacy and applicability of ML algorithms in field-based experiments. The first and foremost challenge is to increase the model's generalizability so that its application is not limited. It would be unwise to believe that our models will be applicable for every contaminated rice-growing site of the world as the models have been trained with a limited set of data and the predictor variables may change on basis of bio-geographical context. However, some points should be taken into consideration as the use of paired soil (rhizosphere soil) and plant samples may be considered for the purpose. The use of huge data set collected from different locations and also, the use of a large number of rice varieties for further studies will enhance the robustness of the model and thereby strengthen the calibration of the model and also its validation.

Acknowledgement:

The authors are grateful to *ICAR – Indian Institute of Water Management, Bhubaneswar, Odisha- 751023, India,* for providing the necessary funds and facilities during the research program. The authors acknowledge the help and cooperation of Mr. Biswanath Halder, the laboratory technical staff. Further, the improvements suggested by the esteemed reviewers are deeply acknowledged.

Conflict of 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.

Author's contributions:

KB conceived the idea of the experiment; SS and JM carried out the experiment and statistical computations; PB, SH and AP contributed in analysis; SS prepared the original draft; KB and JM finally reviewed, edited and compiled the manuscript.

References:

Alexander, D.L.J., Tropsha, A., Winkler A.D. 2015. Beware of R^2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. J Chem Inf Model. 55,1316-1322.

Altman N.S. 1991. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 46 (3), 175–185.

Antoine, J.M., Fung, L.A.H., Grant, C.N. 2017. Assessment of the potential health risks associated with the aluminium, arsenic, cadmium and lead content in selected fruits and vegetables grown in Jamaica. Toxicol. Rep, 4, 181-187.

Awasthi, S., Chauhan, R., Srivastava, S., Tripathi, R.D. 2017. The journey of arsenic from soil to grain in rice. Front. Plant Sci. 8, 1007.

Bakhat H F., Zia Z., Fahad S., Abbas S., Hammad H.M., Shahzad A.N., Abbas F., Alharby H., Shahid M. (2017) Arsenic uptake, accumulation and toxicity in rice plants: Possible remedies for its detoxification: A review. Environ. Sci. Pollut. Res. 24(10), 9142-9158.

Bhattacharyya, K., Sengupta, S., Pari, A., Halder, S., Bhattacharya, P., Pandian, B. J., Chinchmalatpure, A.R. 2021. Characterization and risk assessment of arsenic contamination in soil–plant (vegetable) system and its mitigation through water harvesting and organic amendment. Environ Geochem Health, 1-16.

Biswas, A., Swain, S., Chowdhury, N.R., Joardar, M., Das, A., Mukherjee, M., Roychowdhury, T. 2019. Arsenic contamination in Kolkata metropolitan city: perspective of transportation of agricultural products from arsenic-endemic areas. Environ. Sci. Pollut. Res. 26, 22929–22944.

Bouyoucos, G.J. 1962. Hydrometer method improved for making particle size analyses of soils. Agron J., 54(5),464-465.

Breiman, L. 2001. Random forests. Mach. Learn. 45, 5–32.

Carrijo, D.R., Akbar, N., Reis, A.F.B., Li, C., Gaudin, A.C.M., Parikh, S.J., Green, P.G., Linquist, B.A., 2018. Impacts of variable soil drying in alternate wetting and drying rice systems on yields, grain arsenic content and soil moisture dynamics. Field Crop Res. 222, 101–110.

Carrijo, D.R., Li, C., Parikh, S.J., Linquist, B.A. 2019. Irrigation management for arsenic mitigation in rice grain: Timing and severity of a single soil drying. Sci. Total Environ. 649, 300-307.

Chowdhury, N.R., Das, A., Joardar, M., De, A., Mridha, D., Das, R., Rahman, M.M., Roychowdhury, T. 2020. Flow of arsenic between rice grain and water: Its interaction, accumulation and distribution in different fractions of cooked rice. Sci. Total Environ., 731, 138937.

Chowdhury, N.R., Ghosh, S., Joardar, M., Kar, D., Roychowdhury, T. 2018. Impact of arsenic contaminated groundwater used during domestic scale post harvesting of paddy crop in West Bengal: Arsenic partitioning in raw and parboiled whole grain. *Chemosphere*, 211, 173-184.

Datta, S.P., Subba Rao, A.,Ganeshamurthy, A.N. 1997. Effect of electrolytes coupled with variable stirring on soil pH. J. Indian Soc. Soil Sci.45*,*185–187.

Evelyn, F., Hodges, J.L. 1951. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties (PDF) (Report). USAF School of Aviation Medicine, Randolph Field, Texas.

Fernández-Baca, C.P., McClung, A.M., Edwards, J.D., Codling, E.E., Reddy, V.R., Barnaby, J.Y. 2021. Grain inorganic arsenic content in rice managed through targeted introgressions and irrigation management. Frontiers in plant science, 11, 2284.

Ghosh, K., Das, I., Das, D.K., Sanyal, S.K. 2012. Evaluation of humic and fulvic acid extracts of compost, oilcake, and soils on complex formation with arsenic. Soil Res., CSIRO Publishing. 50, 239–48.

Golui, D., Mazumder, D.G., Sanyal, S.K., Datta, S.P., Ray, P., Patra, P.K., Sarkar, S., Bhattacharya, K. 2017. Safe limit of arsenic in soil in relation to dietary exposure of arsenicosis patient from Malda district West Bengal – A case study. Ecotoxi. Environ. Saf. 144, 227–35.

GRiSP (Global Rice Science Partnership). 2013. Rice almanac, 4th edition. Los Baños (Philippines): International Rice Research Institute. 283 p.

Guha Mazumder, D.N., Deb, D., Biswas, A., Saha, C., Nandy, A., Ganguly, B., Ghose, A., Bhattacharya, K., Majumdar, K.K. 2013. Evaluation of dietary arsenic exposure and its biomarkers: a case study of West Bengal, India. J. Environ. Sci. Health, Part A, 488, 896-904.

Halder, D., Biswas, A., Šlejkovec, Z., Chatterjee, D., Nriagu, J., Jacks, G. 2014. Arsenic species in raw and cooked rice: Implications for human health in rural Bengal. Sci Total Environ. 497-498, 200-208.

Hoffman, A.L., Kemanian, A.R., Forest, C.E. 2018. Analysis of climate signals in the crop yield record of sub-Saharan Africa. Glob Chang Biol, 24(1), 143-157.

Hua, B., Yan, W., Wang, J., Deng, B., Yang, J., 2011. Arsenic accumulation in rice grains: effects of cultivars and water management practices. Environ. Eng. Sci. 28, 591–596.

Jackson M.L. 1973. Soil Chemical Analysis. New Delhi, Prentice Hall.

Johnston, S., Barnard, W.M. 1979. Comparative effectiveness of fourteen solutions for extracting arsenic from four western New York soils. Soil Sci Soc Am J. 43, 304–308.

Knudsen, D., Peterson, G.A.,Pratt, P.F. 1982. Lithium, sodium, and potassium. Methods of soil analysis. Part 2. Chemical and microbiological properties, 225-246.

Kuhn M. 2008. Building Predictive Models in R Using the caret Package. J. Stat. Softw. 28 (5), 1-26.

Kumar S.N., Mishra B.B., Kumar S. and Mandal J. 2021. Organo-arsenic Complexation Studies Explaining the Reduction of Uptake of Arsenic in Wheat Grown with Contaminated Irrigation Water and Organic Amendments. Water Air Soil Pollut., 232:118

Kumarathilaka, P., Seneweera, S., Meharg, A., Bundschuh, J. 2018. Arsenic speciation dynamics in paddy rice soil-water environment: sources, physico-chemical, and biological factors - a review. Water Res. 140, 403–414.

Kumari, P.B., Singh, Y.K., Mandal, J., Shambhavi, S., Sadhu, S.K., Kumar, R., Ghosh, M., Raj, A., Singh, M. 2021. Determination of safe limit for arsenic contaminated irrigation water using solubility free ion activity model (FIAM) and Tobit Regression Model. Chemosphere.270, 128630.

Li, C., Carrijo, D.R., Nakayama, Y., Linquist, B.A., Green, P.G., Parikh, S.J. 2019. Impact of Alternate Wetting and Drying Irrigation on Arsenic Uptake and Speciation in Flooded Rice Systems. Agric Ecosyst Environ, 272, 188-198.

Mandal, J., Golui, D., Datta S.P. 2019a. Assessing equilibria of organo-arsenic complexes and predicting uptake of arsenic by wheat grain from organic matter amended soils. Chemosphere 234, 419–26.

Mandal, J., Golui, D., Raj, A., Ganguly, P. 2019b. Risk Assessment of Arsenic in Wheat and Maize Grown in Organic Matter Amended Soils of Indo-Gangetic Plain of Bihar, India. Soil Sediment Contam28(8), 757-772.

Menon, M., Sarkar, B., Hufton, J., Reynolds, C., Reina, S.V., Young, S. 2020. Do arsenic levels in rice pose a health risk to the UK population?Ecotoxicol. Environ. Saf.197, 110601.

Mondal, D., Banerjee, M., Kundu, M., Banerjee, N., Bhattacharya, U., Giri, A.K., Ganguli, B., Roy, S.S., Polya, D.A. 2010. Comparison of drinking water, raw rice and cooking of rice as arsenic exposure routes in three contrasting areas of West Bengal, India. Environ Geochem Health, 32(6), 463-477.

Muhr, G.R., Datta, N.P., Sankarasubramoney, H., Leley, V.K. and Donahue, R.L. 1965. Soil testing in India. 2nd Ed. USAID Mission to India, New Delhi.

Mukherjee, A., Kundu, M., Basu, B., Sinha, B., Chatterjee, M.,Bairagya, M.D., Singh, U.K., Sarkar, S. 2017. Arsenic load in rice ecosystem and its mitigation through deficit irrigation. J. Environ. Manage, 197, 89-95.

Mukhopadhyay, D., Sanyal, S.K. 2004. Complexation and release isotherm of arsenic in arsenic–humic/fulvic equilibrium study.Aust J Soil Res42, 815–824

Murugaiyan, V., Ali, J., Mahender, A., Aslam, U.M., Jewel, Z.A., Pang, Y., Marfori-Nazarea, C.M., Wu, L.B., Frei, M., Li, Z. 2019. Mapping of genomic regions associated with arsenic toxicity stress in a backcross breeding populations of rice (Oryza sativa L.). *Rice*, *12*(1), 1-14.

Nelder, J.A., Wedderburn, R.W. 1972. Generalized linear models. J R Stat Soc Ser A Stat Soc 135(3), 370-384.

Olsen, S.R., Sommers, L.E. 1982. Phosphorus. p. 403–430. AL Page et al.(ed.) Methods of soil analysis. Part 2. *Agronomy Monographs,* 9. ASA and SSSA, Madison, WI.

Page, A.L., Miller, R.H., Kenny, D.R. 1982. Chemical and microbiologicalproperties. In: Methods of soil analysis, part 2, 2nd ed. Agron.Monogr. 9. ASA and SSSA, Madison, Wisconsin, pp 199-224.

Rahaman, S., Sinha, A.C., Mukhopadhyay, D. 2011. Effect of water regimes and organic matters on transport of arsenic in summer rice (*Oryza sativa* L.). J Environ Sci (China), 23(4), 633-639.

Rahman, M., Islam, M., Hassan, M., Islam, S., Zaman, S. 2015. Impact of water management on the arsenic content of rice grain and cultivated soil in an arsenic contaminated area of Bangladesh. J. Environ. Sci. Nat. Resour.7, 43–46.

Rattan, R.K., Katyal, J.C., Dwivedi, B.S., Sarkar, A.K., Bhattacharyya, T., Tarafdar, J.C. and Kukal, S.S. (Editors).2015. Soil Science:An Introduction. Indian Society of Soil Science, New Delhi.

Reddy, B.G., Reddy, M.S. 1999. Effect of integrated nutrient management on soil available micro nutrients in maize-soybean cropping system. Journal of Res ANGRAU. 27, 24-28.

Ritter, K., Aiken, G.R., Ranville, J.F., Bauer, M., Macalady, D.L. 2006. Evidence for the aquatic binding of arsenate by natural organic matter-suspended Fe(III). Environ. Sci. Technol. 40, 5380–5387.

Saha, D., Basso, B., Robertson, G.P. 2021. Machine learning improves predictions of agricultural nitrous oxide (N_2O) emissions from intensively managed cropping systems. Environ. Res. Lett. 16, 024004.

Sand, S., Bjerselius, R., Busk, L., Eneroth, H., Sanner-Färnstrand, J., Lindqvist, R. 2015. The risk thermometer - a tool for risk comparison. Swedish National Food Agency Report Serial Number 8.

Sanyal, S.K. 2017. A Textbook of Soil Chemistry. Daya Publishing House, A Division of Astral International Pvt. Limited (ISO 9001:2015 Certified Company), New Delhi. ISBN 978-93-5124-865-1 (HB).

Shaheen N., Ahmed, M.K., Islam, M.S., Al-Mamun M.H., Tukun, A.B, Islam, S., Rahim A.T.M.A. 2016. Health risk assessment of trace elements via dietary intake of ‘nonpiscine protein source’ foodstuffs (meat, milk and egg) in Bangladesh.Environ. Sci. Pollut. Res.23(8),7794–7806.

Shrivastava, A., Barla, A., Majumdar, A., Singh, S., Bose, S. 2020. Arsenic mitigation in rice grain loading via alternative irrigation by proposed water management practices. Chemosphere, 238, 124988.

Signes-Pastor, A.J., Mitra, K., Sarkhel, S., Hobbes, M., Burlo, F., De Groot, W.T., Carbonell-Barrachina, A.A. 2008. Arsenic speciation in food and estimation of the dietary intake of inorganic arsenic in a rural village of West Bengal, India. J. Agric. Food Chem.56, 9469– 9474.

Sinha, B., Bhattacharyya, K. 2011. Retention and release isotherm of arsenic in arsenic–humic/fulvic equilibrium study. Biol. Fertil. Soils, 47(7), 815-822.

Sinha, B., Bhattacharyya, K. 2014. Arsenic accumulation and speciation in transplanted autumn rice as influenced by source of irrigation and organic manures. International Journal of Bio-resource and Stress Management. 5(3): 363-368.

Sinha, B., Bhattacharyya, K. 2020. Arsenic speciation in rice and risk assessment of inorganic arsenic from Ghentugachhi village of Chakdaha block, Nadia, West Bengal, India. ORYZA, 57(2), 85-93.

Sinha, B., Bhattacharyya, K., Giri, P.K., Sarkar, S. 2011. Arsenic contamination in sesame and possible mitigation through organic interventions in the lower Gangetic Plain of West Bengal, India. J. Sci. Food Agric., 91(15), 2762-2767.

Song, Y., Liang, J., Lu, J., Zhao, X. 2017. An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing, 251, 26-34.

Sparks D.L., Page A.L., Helmke P.A., Leoppert R.H., Solthanpour P.N., Tabatabai M.A., Johnston C.T., Sumner M.E. 2006. Methods of Soil Analysis. Part 3. Chemical Methods. Madison, Soil Science Society of America, 811–831.

Subbiah, B., Asija, G.L. 1956. Alkaline permanganate method of available nitrogen determination. Curr. Sci., *25,* 259.

Tropsha, A. 2010. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol. Inf. 29, 476−488.

Upadhyay, M.K., Majumdar, A., Barla, A., Bose, S., Srivastava, S. 2019. An assessment of arsenic hazard in groundwater–soil–rice system in two villages of Nadia district, West Bengal, India. Environ Geochem Health.41(6), 2381-2395.

Walkley, A., Black, I.A. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil sci*,* 37(1)*,* 29-38.

Welling, S.H., Refsgaard, H.H., Brockhoff, P.B., Clemmensen, L.H. 2016. Forest floor visualizations of random forests. arXiv preprint arXiv:1605.09196.

Table-1: Arsenic concentration (mg kg^{-1}) and grain yield (t ha^{-1}) of rice under simulated irrigation situations and organic amendment (*pooled data of two year study)*

Interventions	Root As	Shoot As	Grain As	Grain yield
Irrigation situations				
I_1	58.465^{b}	5.491^{b}	0.741^{b}	3.075^{b}
I_2	53.687^{c}	4.359^{c}	0.655^{c}	3.209^{a}
I_3	65.619^{a}	6.724^{a}	0.885^{a}	2.905^{b}
Organic amendments				
F_1	55.772^{c}	4.946^{c}	0.669^{c}	3.118^{ab}
F_2	48.907^{d}	2.586^{d}	0.534^{d}	3.267^{a}
F_3	63.087^{b}	6.324^{b}	0.768^{b}	3.184^{ab}
F_4	69.262^{a}	8.243^{a}	0.844^{a}	2.626^{b}
Interaction				
I_1F_1	54.410^{d}	4.317^{e}	0.662^{ef}	3.194^{b}
I_1F_2	48.417^{e}	2.840^{h}	0.522^{h}	3.233^{ab}
I_1F_3	64.450^{c}	6.537^{d}	0.726^{d}	3.232^{ab}
I_1F_4	68.583^{b}	8.630^{b}	0.755^{c}	2.622^{e}
I_2F_1	50.650^{e}	3.983^{f}	0.416^{i}	3.196^{b}
I_2F_2	43.137^{f}	2.123^{i}	0.318^{j}	3.506^{a}
I_2F_3	57.777^{d}	4.577^{e}	0.559^{g}	3.311^{ab}
I_2F_4	63.183^{c}	6.754^{d}	0.602^{f}	2.826^{de}
I_3F_1	62.257^{c}	6.537^{d}	0.751^{c}	2.966^{d}
I_3F_2	55.167^{d}	3.153^{g}	0.696^{de}	3.061^{c}
I_3F_3	69.033^{b}	7.860^{c}	0.828^{b}	3.010^{c}
I_3F_4	76.020^{a}	9.437^{a}	0.895^{a}	2.589^{e}

Here, I_1= Saturation, I_2= Alternate wetting and drying, I_3= Continuous submergence and F_1=Mustard cake, F_2=Vermicompost, F_3= FYM, F_4=No manure. Means followed by a different letter are significantly different (otherwise statistically at par) at $P < 0.05$ by Duncan's multiple range tests.

Table-2: Dietary risk (TCR and SAMOE) of arsenic through the pooled data of contaminated rice under the applied irrigation management and organic amendments

Treatment	iAs (mg kg^{-1})	TCR	SAMOE
I_1F_1	0.56	4.91x10^{-3}	0.054
I_1F_2	0.44	3.87 x10^{-3}	0.068
I_1F_3	0.61	5.38 x10^{-3}	0.049
I_1F_4	0.63	5.59 x10^{-3}	0.047
I_2F_1	0.35	3.08 x10^{-3}	0.086
I_2F_2	**0.27**	**2.36 x10^{-3}**	**0.112**
I_2F_3	0.47	4.14 x10^{-3}	0.064
I_2F_4	0.51	4.46 x10^{-3}	0.059
I_3F_1	0.63	5.57 x10^{-3}	0.048
I_3F_2	0.58	5.16 x10^{-3}	0.051
I_3F_3	0.70	6.14 x10^{-3}	0.043
I_3F_4	**0.75**	**6.64 x10^{-3}**	**0.040**

Here, I_1= Saturation, I_2= Alternate wetting and drying, I_3= Continuous submergence and F_1=Mustard cake, F_2=Vermicompost, F_3= FYM, F_4=No manure. ***Inorganic arsenic (****iAs) is obtained by multiplying total arsenic (as in Table-1, pooled data) by 0.866, referring to Sinha and Bhattacharyya (2020), who opined~86.6% of total As in IET-4786 is inorganic.* ***TCR,*** *indicating target cancer risk of the carcinogenic As is computed based on exposure frequency to arsenic (365 days) over an exposure duration (70 yrs) accruing averaged carcinogenic exposure time (365days*70yrs). Further it includes 400 g rice consumption daily by an individual of 68 kg and oral daily cancer slope for As (1.5 mg/kg). Any value above 10^{-4} is detrimental for health.* ***SAMOE*** *(Severity Adjusted Margin of Exposure) has its expression using the assumptions that 3.0 µg kg^{-1} bodyweight^{-1} is the threshold daily intake of As, a value of 10 for assessment factors, 1/10 of Benchmark response and a severity factor of 100; values below 0.1 are risky for human.*

Table-3: Effect of organic amendments and irrigation management on post harvest soil properties under cultivation of rice (pooled for two years)

	pH	**OC (%)**	**Available P (mg kg^{-1})**	**Available As (mg kg^{-1})**
I_1F_1	6.88^{g}	0.51^{de}	14.05^{f}	6.557^{f}
I_1F_2	6.84^{h}	0.57^{b}	15.55^{c}	6.013^{h}
I_1F_3	6.92^{f}	0.53^{c}	14.59^{d}	7.027^{e}
I_1F_4	6.93^{e}	0.51^{de}	14.36^{de}	7.697^{d}
I_2F_1	6.94^{e}	0.52^{d}	14.50^{de}	6.157^{gh}
I_2F_2	6.81^{i}	0.59^{a}	16.45^{a}	5.623^{i}
I_2F_3	6.96^{d}	0.52^{d}	14.32^{e}	6.667^{f}
I_2F_4	6.97^{c}	0.51^{de}	14.32^{e}	7.070^{e}
I_3F_1	6.97^{c}	0.49^{e}	14.18^{ef}	7.987^{c}
I_3F_2	6.92^{f}	0.57^{b}	16.23^{b}	6.293^{g}
I_3F_3	6.99^{b}	0.53^{c}	16.00^{b}	8.820^{b}
I_3F_4	7.01^{a}	0.47^{f}	13.68^{g}	9.430^{a}

Here, I_1= Saturation, I_2= Alternate wetting and drying, I_3= Continuous submergence and F_1=Mustard cake, F_2=Vermicompost, F_3= FYM, F_4=No manure. Means followed by a different letter are significantly different (otherwise statistically at par) at $P < 0.05$ by Duncan's multiple range tests.

Table -4: Comparison between the performance matrices of the models (n=36)

Machine Learning Algorithms	RMSE	MAE	R^2
Random Forest (RF)	0.065	0.055	0.86
k-Nearest Neighbour (KNN)	0.066	0.056	0.88
Generalized Linear Model (GLM)	0.086	0.070	0.77

Here, coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) have been estimated to compare the model performances

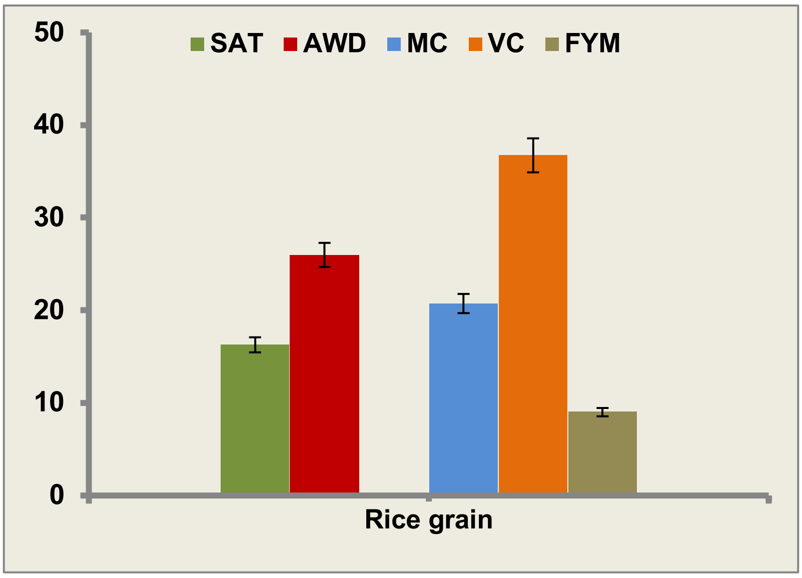


Fig-1. Percent reduction in arsenic recoveries of rice grain through irrigation management and organic amendment (pooled data) in comparison to farmer's practice of continuous submergence and no manure situation

Here, SAT=Saturation, AWD= Alternate wetting and drying, MC=Mustard cake, VC=Vermicompost, FYM= Farm Yard Manure

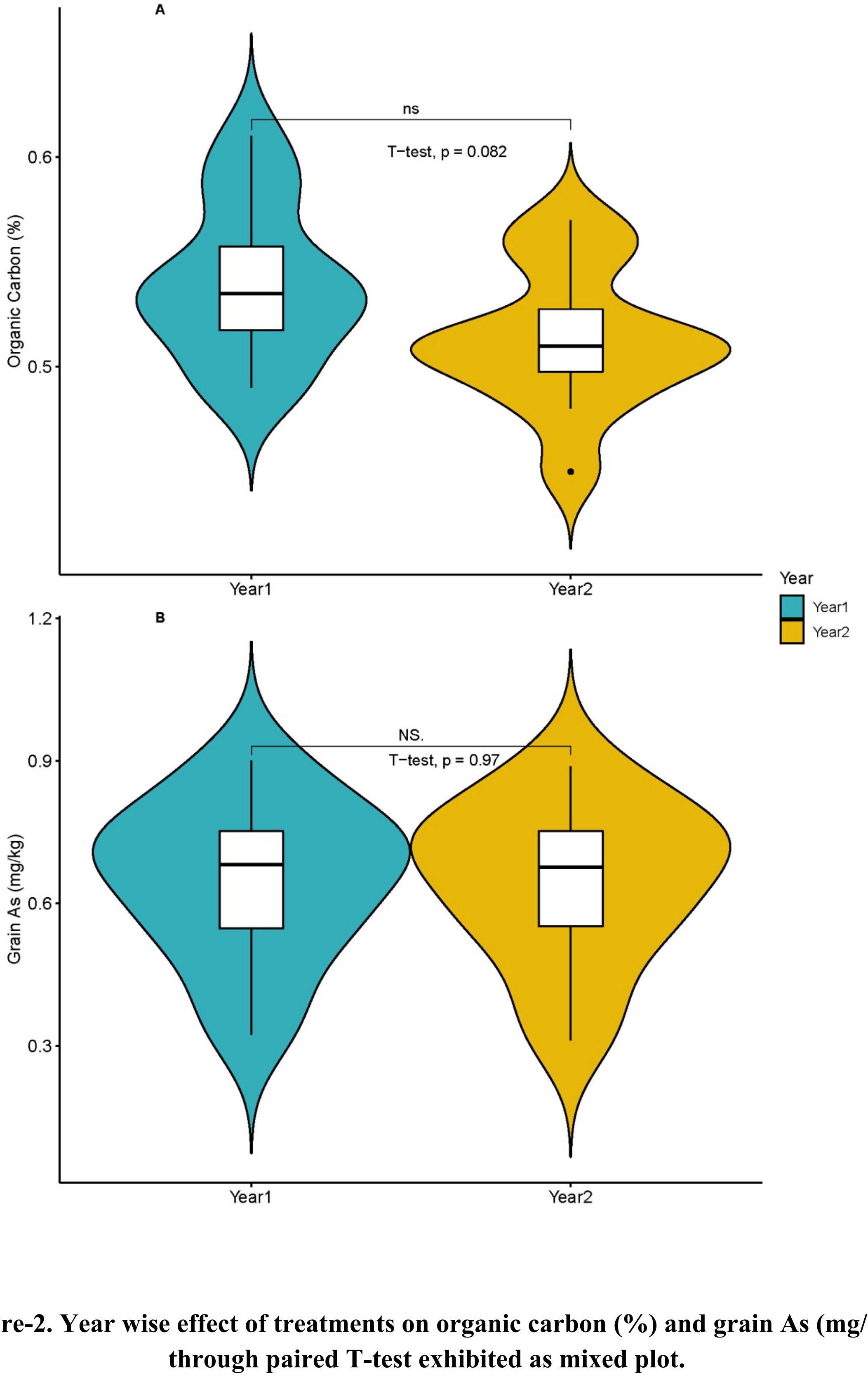


Figure-2. Year wise effect of treatments on organic carbon (%) and grain As (mg/kg) through paired T-test exhibited as mixed plot.

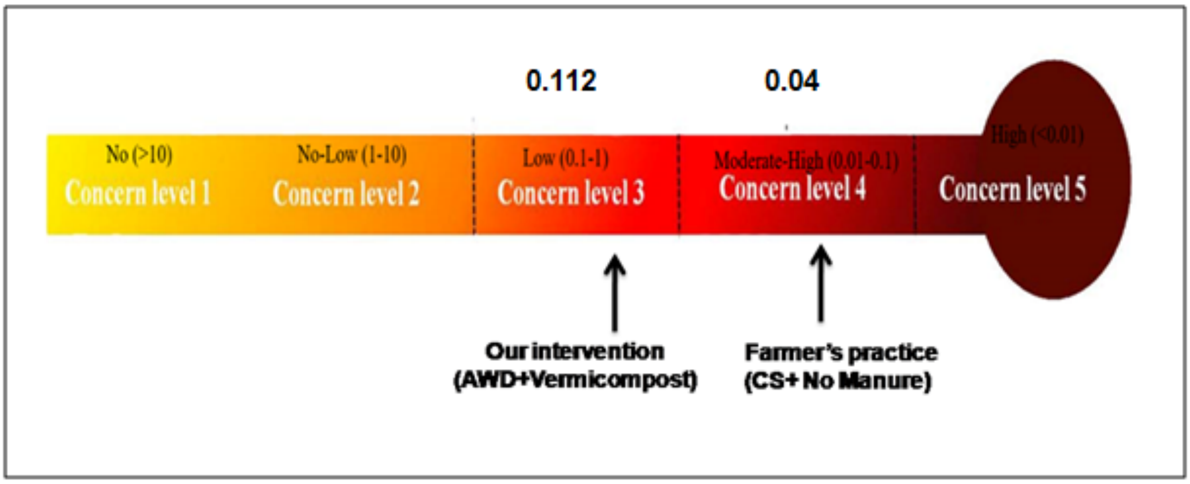


Fig-3. Risk thermometer scale showing the class of arsenic toxicity through intake of rice cultivated under different water and organic management regimes

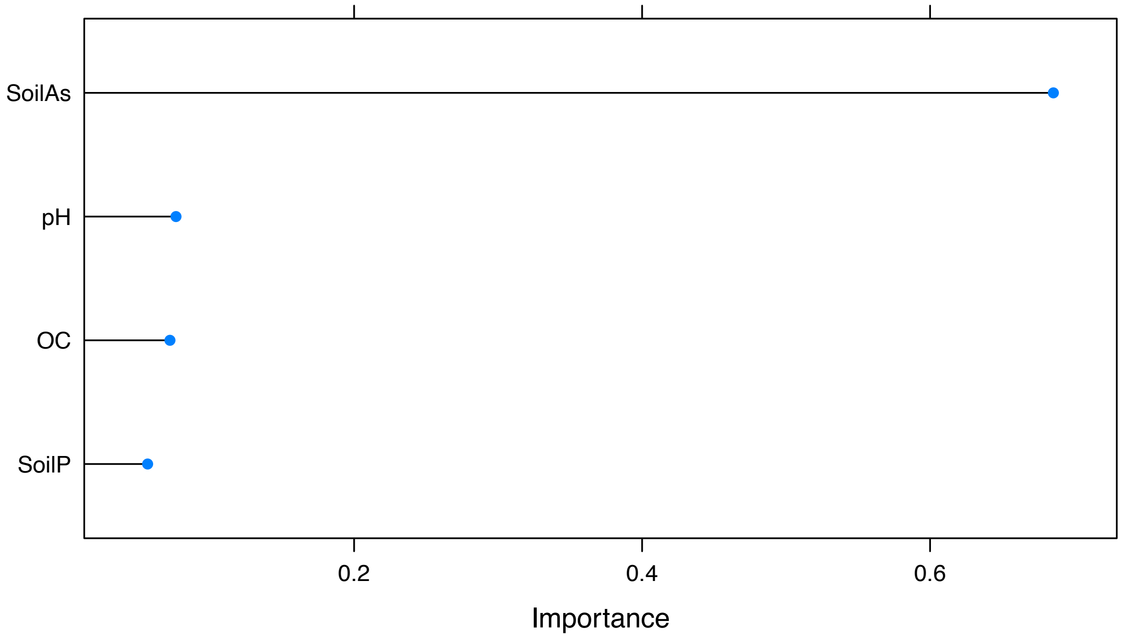


Fig-4. Variable importance plot with Random Forest algorithm

Supplementary:

Journal: Agriculture, Ecosystems and Environment

Deficit irrigation and organic amendments can reduce dietary arsenic risk from rice: introducing machine learning-based prediction models from field data

Sudip Sengupta[a], Kallol Bhattacharyya[a], Jajati Mandal[b,c*], Parijat Bhattacharya[a], Sanjay Halder[a] and Arnab Pari[a]

[a] *Department of Agricultural Chemistry and Soil Science, Faculty of Agriculture, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur 741252, Nadia, West Bengal, India*

[b] *Department of Soil Science and Agricultural Chemistry, Bihar Agricultural University, Sabour, Bhagalpur 813210, Bihar, India.*

[c] *School of Science, Engineering and Environment, University of Salford, Manchester, M5 4WT, United Kingdom*

[*]***Corresponding author:*** jajati.bckv@gmail.com

List of Supplementary materials:

Table S1: Ratings of fertility status of soil

Table S2: Target cancer risk (TCR) of arsenic through the contaminated rice grain

Table S3: SAMOE for As toxicity through the contaminated rice grain

Fig-S1. Experimental design of the study on rice for both the study years under employed irrigation management and organic interventions.

Table S1: Ratings of fertility status of soil (Muhr et al., 1965; Rattan et al.,2015)

Nutrient (kg ha^{-1})	**Fertility Rating**		
	Low	Medium/ Moderate	High
Nitrogen	≤ 280	281-560	> 560
Phosphorus	≤ 10	11-25	> 25
Potassium	≤ 120	121-280	> 280

Table S2: Target cancer risk (TCR) of arsenic through the contaminated rice grain

Treatment	Efr (days)	Ed (years)	Fir (g/day)	C (mg/kg)	CPSo (mg/kg day^{-1})	BWa (kg)	ATc (days)	TCR
I_1F_1	365	70	400	0.56	1.5	68	25550	4.91x10^{-3}
I_1F_2	365	70	400	0.44	1.5	68	25550	3.87 x10^{-3}
I_1F_3	365	70	400	0.61	1.5	68	25550	5.38 x10^{-3}
I_1F_4	365	70	400	0.63	1.5	68	25550	5.59 x10^{-3}
I_2F_1	365	70	400	0.35	1.5	68	25550	3.08 x10^{-3}
I_2F_2	**365**	**70**	**400**	**0.27**	**1.5**	**68**	**25550**	**2.36 x10^{-3}**
I_2F_3	365	70	400	0.47	1.5	68	25550	4.14 x10^{-3}
I_2F_4	365	70	400	0.51	1.5	68	25550	4.46 x10^{-3}
I_3F_1	365	70	400	0.63	1.5	68	25550	5.57 x10^{-3}
I_3F_2	365	70	400	0.58	1.5	68	25550	5.16 x10^{-3}
I_3F_3	365	70	400	0.70	1.5	68	25550	6.14 x10^{-3}
I_3F_4	**365**	**70**	**400**	**0.75**	**1.5**	**68**	**25550**	**6.64 x10^{-3}**

Here, I_1= Saturation, I_2= Alternate wetting and drying, I_3= Continuous submergence and F_1=Mustard cake, F_2=Vermicompost, F_3= FYM, F_4=No manure.

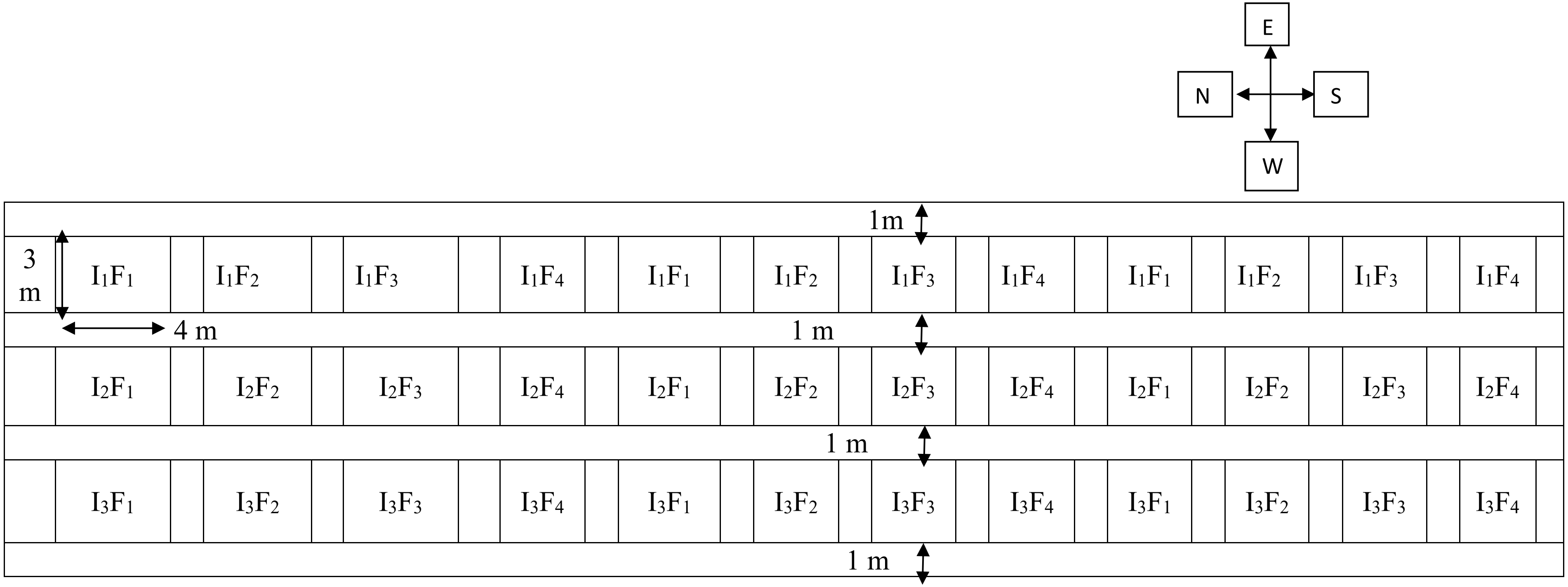
Inorganic As obtained by multiplying total As by 0.866, referring to Sinha and Bhattacharyya (2020), who opined~86.6% of total As in IET-4786 is inorganic

Table- S3: SAMOE for As toxicity through the contaminated rice grain

Treatment	TDI	AF_{BMR}	AF	SF	E	SAMOE
I_1F_1	3	0.1	10	100	0.56	0.054
I_1F_2	3	0.1	10	100	0.44	0.068
I_1F_3	3	0.1	10	100	0.61	0.049
I_1F_4	3	0.1	10	100	0.63	0.047
I_2F_1	3	0.1	10	100	0.35	0.086
I_2F_2	**3**	**0.1**	**10**	**100**	**0.27**	**0.112**
I_2F_3	3	0.1	10	100	0.47	0.064
I_2F_4	3	0.1	10	100	0.51	0.059
I_3F_1	3	0.1	10	100	0.63	0.048
I_3F_2	3	0.1	10	100	0.58	0.051
I_3F_3	3	0.1	10	100	0.70	0.043
I_3F_4	**3**	**0.1**	**10**	**100**	**0.75**	**0.040**

Here, I_1*= Saturation,* I_2*= Alternate wetting and drying,* I_3*= Continuous submergence and* F_1*=Mustard cake,* F_2*=Vermicompost,* F_3*= FYM,* F_4*=No manure.*

Inorganic As obtained by multiplying total As by 0.866, referring to Sinha and Bhattacharyya (2020), who opined~86.6% of total As in IET-4786 is inorganic



Here, I_1*= Saturation,* I_2*= Alternate wetting and drying,* I_3*= Continuous submergence and* F_1*=Mustard cake,* F_2*=Vermicompost,* F_3*= FYM,* F_4*=No manure. The experimental layout comprised of 36 plots, each of 3m × 4m in size, replicated thrice and laid in strip plot design*

Fig-S1. Experimental design of the study on rice for both the study years under employed irrigation management and organic interventions.