Supplementary Material for

Exploring the key soil parameters relevant to arsenic and cadmium accumulation in rice grain in southern China

Xu Fang[1], Anna Muntwyler[1], Pascal Schneider[1], Iso Christl[1,*], Peng Wang[2], Fang-Jie Zhao[2], Ruben Kretzschmar[1]

1. Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Switzerland

2. College of Resources and Environmental Sciences, Nanjing Agricultural University, China

* Corresponding author: iso.christl@env.ethz.ch

1. Methods

1.1. Soil characterization

To determine soil total elemental concentrations, 1 g of dry soil was extracted by 10 mL aqua regia (molar ratio of HNO_3 to HCl was 1:3) on a heating plate at 120 °C for 90 min before topping up the digestion tube to 50 mL by deionized water and filtrating (0.45 µm, nylon filter) the soil digest solution for ICP-MS measurement. Soil total organic carbon (TOC) was determined as total carbon (CHNS-923, LECO) for soils with pH_d < 6.35 and as the difference between total carbon and inorganic carbon (TOC-L, SHIMADZU) for soils with pH_d > 6.35 using sieved (< 2 mm) and milled dry soils. Soil texture (fractions of sand, silt and clay) was determined by commercial lab (Sol-Conseil, Switzerland) using sieved (< 2 mm) dry soils.

1.2. As speciation of brown rice

For As speciation analysis, triplicates of 1.5 g milled rice grains and 15 mL 0.28 M HNO_3 were mixed in 50 mL polypropylene tube and heated to 90 °C in heating blocks (SCP Science, DigiPREP MS) for 1.5 hours. At the end of extraction, tubes were immediately transferred into a water bath (23°C) for cooling down to minimize further redox reactions in samples. After centrifugation at 3300 *g* for 10 minutes, the supernatants were filtered (0.45µm, PTFE) and stored at 4 °C in the dark prior to analysis. Arsenite, dimethylarsenate, monomethylarsenate and arsenate were determined by HPLC-ICP-MS using an anion exchange column (Hamilton PRP-X100) and malonic acid (buffered pH 5.2, adjusted with $NH_3 \cdot H_2O$) as the mobile phase. The concentrations of arsenite and arsenate were summed up to calculate inorganic arsenic (iAs). Due to limited amounts of milled rice grains for few samples, As speciation was carried out for 26 samples only.

1.3. Selection of best multivariate linear regression models

For k-variable (k = 1, 2, 3) models, models were constructed containing all possible combinations of k variables, resulting in $\binom{k}{n}$ alternative models for subsequent selection. For each of these models, the adjusted R-squared from robust linear regression ($R^2_{adj\ rob}$, denoted as R^2_{adj} in the following) of each model was also calculated. and the averaged cross-validated root-mean-squared-error (CV-RMSE) after 100 times repetitions of 5-fold cross validation were calculated. Specifically, the 31 samples were randomly and evenly split into five subsets where four subsets was re-combined as training set to fit one k-variable linear regression

model and the remaining one subset was used as test set to validate the model through the calculation of root-mean-squared-error (RMSE). Each of the five subsets was used as test set once (what "cross-validation" refers to) and the above calculation was repeated to produce five RMSE values which were averaged as the CV-RMSE of this sample split. By repeating the random sample splitting and subsequent CV-RMSE calculation 100 times, the averaged CV-RMSE was calculated. To start model selection, all k-variable models were ranked by R^2_{adj}. The top 100 (highest R^2_{adj}) models (or all models if number of the latter is smaller) were used to calculate their CV-RMSE. Models with CV-RMSE lower than the average CV-RMSE among those 100 models were further selected and ranked again by their R^2_{adj}. Finally, models were categorized according to whether the target element (Cd or As) was present in any of the model parameters. Models without the target element in any of their explanatory variable(s) were ranked after those with target element in their explanatory variable(s). In this way, the top 5 models were selected for further interpretation. The aforementioned selection was conducted by using an R function developed based on the "bestlinearmodels" function by Simmler et al in R Studio (Version 1.4.1106).

2. Results

Table S1. Detection limits and quality controls (recoveries for certified rice and soil samples).

	Rice total (RT)		Soil total (ST)		AscCit extractable (AscCit)		Incubation dissolved (Diss)		Dry soil exchangeable (DSE)		Moist soil exchangeable (SE)	
	DL	**QC**	**DL**	**QC**	**DL**	**QC**	**DL**	**QC**	**DL**	**QC**	**DL**	**QC**
	µg/kg	%	mg/kg	%	mg/kg	%	µg/kg	%	µg/kg	%	µg/kg	%
As	6.5	99	0.01	104	0.21	92	1.1	107	2.6	93	3.3	106
Cd	1.9	104	0.001	96	0.06	91	0.09	97	1.3	99	0.27	102
Fe	11	91	0.10	101	3.3	103	11	99	5.0	105	39	115
Mn	6.4	100	0.20	105	0.18	103	0.10	96	70	102	2.3	110
S	647	109	3.9	77	5.6	89	17	101	181	102	1605	–
Si	333	–	–	–	–	–	28	–	–	–	333	–
Zn	10	91	0.30	96	3.0	98	1.2	98	27	95	12	102

Note: Certified reference material for each analysis is as follow: NIST 1568b rice powder for rice total; In-house reference soil for soil total; diluted mixed standard solution independent from calibration standards for other analyses.

Table S2. Top five k-variable models for grain As built from soil parameter set 1.

			Grain As models: Parameter set 1					
			lg(As_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable			0.078 lg(As_ST)	+ 2.138			0.005	0.181
	1	weight (%)	100					
		p value	0.290					
			-0.200 lg(Mg_ST)	+ 2.980			0.113	0.171
	2	weight (%)	100					
		p value	0.033					
			-0.237 lg(clay)	+ 2.590			0.032	0.178
	3	weight (%)	100					
		p value	0.164					
			-0.113 lg(Ni_ST)	+ 2.399			0.031	0.180
	4	weight (%)	100					
		p value	0.161					
			-0.090 lg(Cu_ST)	+ 2.383			0.003	0.184
	5	weight (%)	100					
		p value	0.294					
2-variable			0.318 lg(As_ST)	- 0.380 lg(Cu_ST)	+ 2.362		0.334	0.163
	1	weight (%)	50	50				
		p value	0.001	<0.001				
			0.161 lg(As_ST)	- 0.202 lg(Ni_ST)	+ 2.289		0.137	0.176
	2	weight (%)	43	57				
		p value	0.038	0.024				
			0.267 lg(As_ST)	- 0.449 lg(Fe_ST)	+ 3.900		0.126	0.175
	3	weight (%)	59	41				
		p value	0.020	0.039				
			0.122 lg(As_ST)	- 0.332 lg(clay)	+ 2.555		0.092	0.174
	4	weight (%)	41	59				
		p value	0.095	0.058				
			0.086 lg(As_ST)	- 0.207 lg(Mg_ST)	+ 2.883		0.128	0.171
	5	weight (%)	22	78				
		p value	0.209	0.028				
3-variable			0.449 lg(As_ST)	- 0.344 lg(Cu_ST)	- 0.379 lg(Fe_ST)	+ 3.833	0.440	0.157
	1	weight (%)	51	36	13			
		p value	<0.001	0.001	0.048			
			0.349 lg(As_ST)	- 0.397 lg(Cu_ST)	- 0.085 lg(Ca_ST)	+ 2.634	0.360	0.164
	2	weight (%)	48	48	4			
		p value	<0.001	<0.001	0.226			
			0.358 lg(As_ST)	- 0.369 lg(Cu_ST)	- 0.126 lg(Mn_ST)	+ 2.602	0.358	0.165
	3	weight (%)	49	46	5			
		p value	<0.001	0.001	0.277			
			0.336 lg(As_ST)	- 0.370 lg(Cu_ST)	+ 0.128 lg(sand)	+ 2.143	0.350	0.164
	4	weight (%)	49	45	6			
		p value	<0.001	0.001	0.307			
			0.328 lg(As_ST)	- 0.350 lg(Cu_ST)	- 0.210 lg(clay)	+ 2.603	0.346	0.161
	5	weight (%)	48	38	14			
		p value	<0.001	0.002	0.210			

Note: P values for intercept were not shown. (same note for Table S3–S9).

Table S3. Top five k-variable models for grain As built from soil parameter set 2.

			Grain As models: Parameter set 2					
			lg(As_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable			0.112 lg(As_AscCit)	+ 2.183			0.021	0.181
	1	weight (%)	100					
		p value	0.211					
			-0.237 lg(clay)	+ 2.590			0.032	0.178
	2	weight (%)	100					
		p value	0.164					
			-0.171 lg(TOC)	+ 2.303			-0.007	0.185
	3	weight (%)	100					
		p value	0.372					
			-0.124 lg(Al_AscCit)	+ 2.539			-0.015	0.183
	4	weight (%)	100					
		p value	0.451					
			0.050 lg(P_AscCit)	+ 2.156			-0.015	0.182
	5	weight (%)	100					
		p value	0.460					
2-variable			0.243 lg(As_AscCit)	- 0.236 lg(S_AscCit)	+ 2.556		0.149	0.187
	1	weight (%)	61	39				
		p value	0.024	0.045				
			0.141 lg(As_AscCit)	- 0.290 lg(clay)	+ 2.585		0.095	0.177
	2	weight (%)	47	53				
		p value	0.096	0.081				
			0.135 lg(As_AscCit)	- 0.221 lg(TOC)	+ 2.241		0.040	0.185
	3	weight (%)	63	37				
		p value	0.132	0.243				
			0.152 lg(As_AscCit)	- 0.102 lg(Co_AscCit)	+ 2.152		0.036	0.186
	4	weight (%)	71	29				
		p value	0.110	0.244				
			0.142 lg(As_AscCit)	+ 0.156 lg(sand)	+ 1.946		0.033	0.181
	5	weight (%)	68	32				
		p value	0.126	0.253				
3-variable			0.297 lg(As_AscCit)	- 0.262 lg(S_AscCit)	+ 0.196 lg(sand)	+ 2.300	0.205	0.183
	1	weight (%)	53	32	15			
		p value	0.007	0.022	0.153			
			0.270 lg(As_AscCit)	- 0.221 lg(S_AscCit)	- 0.103 lg(Co_AscCit)	+ 2.505	0.183	0.186
	2	weight (%)	56	32	13			
		p value	0.012	0.058	0.239			
			0.122 lg(As_AscCit)	+ 0.124 lg(Mn_AscCit)	- 0.487 lg(clay)	+ 2.660	0.170	0.180
	3	weight (%)	26	21	53			
		p value	0.118	0.086	0.012			
			0.185 lg(As_AscCit)	+ 0.277 lg(Mn_AscCit)	- 0.415 lg(Co_AscCit)	+ 1.624	0.158	0.175
	4	weight (%)	36	27	38			
		p value	0.054	0.045	0.019			
			0.142 lg(As_AscCit)	- 0.162 lg(Al_AscCit)	- 0.311 lg(clay)	+ 2.990	0.097	0.178
	5	weight (%)	38	14	48			
		p value	0.093	0.287	0.063			

Table S4. Top five k-variable models for grain As built from soil parameter set 3.

			Grain As models: Parameter set 3					
			lg(As_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable			0.098 lg(As_Diss)	+ 2.009			0.020	0.185
	1	weight (%)	100					
		p value	0.219					
			-0.153 lg(S_Diss)	+ 2.793			0.113	0.191
	2	weight (%)	100					
		p value	0.038					
			-0.237 lg(clay)	+ 2.590			0.032	0.178
	3	weight (%)	100					
		p value	0.164					
			-0.132 lg(Co_Diss)	+ 2.384			0.030	0.187
	4	weight (%)	100					
		p value	0.175					
			-0.170 lg(TOC)	+ 2.302			-0.007	0.185
	5	weight (%)	100					
		p value	0.372					
2-variable			0.141 lg(As_Diss)	- 0.167 lg(S_Diss)	+ 2.505		0.238	0.183
	1	weight (%)	34	66				
		p value	0.063	0.018				
			0.130 lg(As_Diss)	- 0.167 lg(Co_Diss)	+ 2.103		0.150	0.186
	2	weight (%)	48	52				
		p value	0.102	0.082				
			0.112 lg(As_Diss)	- 0.270 lg(clay)	+ 2.364		0.062	0.184
	3	weight (%)	45	55				
		p value	0.149	0.129				
			0.114 lg(As_Diss)	- 0.207 lg(TOC)	+ 2.037		0.031	0.188
	4	weight (%)	64	36				
		p value	0.153	0.300				
			0.107 lg(As_Diss)	- 0.066 lg(Fe_Diss)	+ 2.289		0.027	0.188
	5	weight (%)	65	35				
		p value	0.185	0.361				
3-variable			0.162 lg(As_Diss)	- 0.188 lg(S_Diss)	- 0.331 lg(clay)	+ 3.005	0.318	0.179
	1	weight (%)	29	51	20			
		p value	0.025	0.006	0.061			
			0.157 lg(As_Diss)	- 0.173 lg(S_Diss)	+ 0.184 lg(sand)	+ 2.231	0.263	0.179
	2	weight (%)	32	58	10			
		p value	0.038	0.013	0.181			
			0.116 lg(As_Diss)	- 0.191 lg(S_Diss)	+ 0.077 lg(Ni_Diss)	+ 2.577	0.234	0.200
	3	weight (%)	13	73	15			
		p value	0.147	0.011	0.480			
			0.147 lg(As_Diss)	- 0.156 lg(S_Diss)	- 0.117 lg(TOC)	+ 2.491	0.218	0.190
	4	weight (%)	35	58	6			
		p value	0.059	0.030	0.567			
			0.152 lg(As_Diss)	- 0.157 lg(S_Diss)	- 0.054 lg(OC_Diss)	+ 2.559	0.212	0.197
	5	weight (%)	35	60	5			
		p value	0.065	0.034	0.738			

Table S5. Top five k-variable models for grain As built from combined soil parameter set.

			Grain As models: Combined parameter set					
			lg(As_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable			0.112 lg(As_AscCit)	+ 2.183			0.021	0.181
	1	weight (%)	100					
		p value	0.211					
			0.098 lg(As_Diss)	+ 2.009			0.020	0.185
	2	weight (%)	100					
		p value	0.219					
			0.078 lg(As_ST)	+ 2.138			0.005	0.181
	3	weight (%)	100					
		p value	0.290					
			-0.200 lg(Mg_ST)	+ 2.980			0.113	0.171
	4	weight (%)	100					
		p value	0.033					
			-0.153 lg(S_Diss)	+ 2.793			0.113	0.191
	5	weight (%)	100					
		p value	0.038					
2-variable			0.277 lg(As_AscCit)	- 0.240 lg(S_Diss)	+ 2.945		0.364	0.167
	1	weight (%)	39	61				
		p value	0.004	<0.001				
			0.318 lg(As_ST)	- 0.380 lg(Cu_ST)	+ 2.362		0.334	0.163
	2	weight (%)	50	50				
		p value	0.001	<0.001				
			0.159 lg(As_ST)	- 0.191 lg(S_Diss)	+ 2.709		0.244	0.181
	3	weight (%)	32.00	68.00				
		p value	0.042	0.007				
			0.141 lg(As_Diss)	- 0.167 lg(S_Diss)	+ 2.505		0.238	0.183
	4	weight (%)	34.00	66.00				
		p value	0.063	0.018				
			0.147 lg(As_Diss)	- 0.245 lg(Mg_ST)	+ 2.786		0.203	0.168
	5	weight (%)	32	68				
		p value	0.050	0.015				
3-variable			0.332 lg(As_AscCit)	- 0.271 lg(S_Diss)	- 0.423 lg(clay)	+ 3.642	0.508	0.153
	1	weight (%)	35	47	18			
		p value	<0.001	<0.001	0.006			
			0.341 lg(As_AscCit)	- 0.267 lg(S_Diss)	- 0.200 lg(Ni_Diss)	+ 3.275	0.489	0.154
	2	weight (%)	38	45	17			
		p value	<0.001	<0.001	0.010			
			0.335 lg(As_AscCit)	- 0.265 lg(S_Diss)	+ 0.287 lg(sand)	+ 2.606	0.459	0.151
	3	weight (%)	37	53	10			
		p value	<0.001	<0.001	0.023			
			0.449 lg(As_ST)	- 0.344 lg(Cu_ST)	- 0.379 lg(Fe_ST)	+ 3.833	0.440	0.157
	4	weight (%)	51	36	13			
		p value	<0.001	0.001	0.048			
			0.275 lg(As_AscCit)	- 0.209 lg(S_Diss)	- 0.187 lg(Mg_ST)	+ 3.520	0.429	0.159
	5	weight (%)	35.00	41.00	23.00			
		p value	0.003	0.002	0.046			

Table S6. Top five k-variable models for grain Cd built from soil parameter set 1.

			Grain Cd models: Parameter set 1					
			lg(Cd_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable			0.116 lg(Cd_ST)	+ 2.681			-0.025	0.623
	1	weight (%)	100					
		p value	0.622					
			-0.366 pH_d	+ 4.606			0.251	0.551
	2	weight (%)	100					
		p value	0.003					
			-0.527 lg(Ca_ST)	+ 4.466			0.086	0.620
	3	weight (%)	100					
		p value	0.070					
			0.686 lg(K_ST)	- 0.147			0.072	0.598
	4	weight (%)	100					
		p value	0.081					
			0.510 lg(Mg_ST)	+ 0.833			0.037	0.594
	5	weight (%)	100					
		p value	0.153					
2-variable			0.181 lg(Cd_ST)	- 0.377 pH_d	+ 4.648		0.249	0.562
	1	weight (%)	5	95				
		p value	0.395	0.002				
			0.159 lg(Cd_ST)	+ 0.715 lg(K_ST)	- 0.291		0.054	0.611
	2	weight (%)	10	90				
		p value	0.489	0.079				
			0.113 lg(Cd_ST)	+0.507 lg(Mg_ST)	+ 0.824		0.011	0.608
	3	weight (%)	11	89				
		p value	0.619	0.166				
			0.481 lg(Mg_ST)	- 0.360 pH_d	+ 2.813		0.283	0.544
	4	weight (%)	19	81				
		p value	0.148	0.003				
			0.345 lg(Cu_ST)	- 0.370 pH_d	+ 4.105		0.270	0.547
	5	weight (%)	11	89				
		p value	0.233	0.002				
3-variable			0.184 lg(Cd_ST)	- 0.372 pH_d	+ 0.478 lg(Mg_ST)	+ 2.871	0.284	0.554
	1	weight (%)	5	78	18			
		p value	0.379	0.002	0.150			
			0.805 lg(Cd_ST)	- 0.405 pH_d	- 0.720 lg(Zn_ST)	+ 6.349	0.274	0.550
	2	weight (%)	11	83	6			
		p value	0.098	0.001	0.143			
			0.178 lg(Cd_ST)	- 0.372 pH_d	+ 0.985 lg(Co_ST)	+ 3.388	0.250	0.573
	3	weight (%)	5	88	7			
		p value	0.406	0.003	0.359			
			0.084 lg(Cd_ST)	- 0.374 pH_d	+ 0.293 lg(Cu_ST)	+ 4.200	0.247	0.566
	4	weight (%)	2	89	9			
		p value	0.725	0.003	0.368			
			0.146 lg(Cd_ST)	- 0.409 pH_d	+ 0.268lg(Ni_ST)	+ 4.470	0.246	0.561
	5	weight (%)	4	92	4			
		p value	0.502	0.002	0.386			

Table S7. Top five k-variable models for grain Cd built from soil parameter set 2.

			Grain Cd models: Parameter set 2					
			lg(Cd_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable	1		0.472 lg(Cd_DSE)	+ 1.793			0.379	0.481
		weight (%)	100					
		p value	<0.001					
	2		0.489 lg(Co_DSE)	+ 1.871			0.278	0.535
		weight (%)	100					
		p value	0.002					
	3		-0.366 pH_d	+ 4.606			0.251	0.551
		weight (%)	100					
		p value	0.003					
	4		0.424 lg(Ni_DSE)	+ 1.924			0.186	0.572
		weight (%)	100					
		p value	0.010					
	5		0.369 lg(Mn_DSE)	+ 1.229			0.106	0.608
		weight (%)	100					
		p value	0.050					
2-variable	1		0.733 lg(Cd_DSE)	- 0.327 lg(Zn_DSE)	+ 2.349		0.479	0.464
		weight (%)	85	15				
		p value	<0.001	0.011				
	2		0.677 lg(Cd_DSE)	- 0.393 lg(Mn_DSE)	+ 2.951		0.419	0.511
		weight (%)	79	21				
		p value	<0.001	0.076				
	3		0.530 lg(Cd_DSE)	- 0.379 lg(S_DSE)	+ 3.449		0.404	0.511
		weight (%)	95	5				
		p value	<0.001	0.161				
	4		0.697 lg(Cd_DSE)	-0.332 lg(Ni_DSE)	+ 1.945		0.387	0.529
		weight (%)	72	28				
		p value	0.003	0.216				
	5		0.478 lg(Cd_DSE)	+ 0.405 lg(P_DSE)	+ 0.686		0.386	0.496
		weight (%)	95	5				
		p value	<0.001	0.206				
3-variable	1		0.712 lg(Cd_DSE)	- 0.325 lg(Zn_DSE)	+ 0.155 lg(Cu_DSE)	+ 2.168	0.492	0.461
		weight (%)	79	14	7			
		p value	<0.001	0.010	0.246			
	2		0.814 lg(Cd_DSE)	- 0.287 lg(Zn_DSE)	- 0.230 lg(Mn_DSE)	+ 2.970	0.486	0.453
		weight (%)	74	11	15			
		p value	<0.001	0.029	0.289			
	3		0.567 lg(Cd_DSE)	- 0.390 lg(Zn_DSE)	+ 0.312 lg(Co_DSE)	+ 2.356	0.485	0.458
		weight (%)	51	18	31			
		p value	0.007	0.007	0.227			
	4		0.637 lg(Cd_DSE)	- 0.338 lg(Zn_DSE)	- 0.159 pH_d	+ 3.403	0.484	0.470
		weight (%)	60	13	27			
		p value	<0.001	0.011	0.216			
	5		0.715 lg(Cd_DSE)	- 0.324 lg(Zn_DSE)	+ 0.270 lg(sand)	+ 1.998	0.467	0.455
		weight (%)	81	15	4			
		p value	<0.001	0.013	0.449			

Table S8. Top five k-variable models for grain Cd built from soil parameter set 3.

			Grain Cd models: Parameter set 3					
			lg(Cd_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable			0.370 lg(Cd_SE)	+ 2.149			0.469	0.487
	1	weight (%)	100					
		p value	<0.001					
			-0.430 pH_m	+ 5.006			0.299	0.533
	2	weight (%)	100					
		p value	<0.001					
			-0.366 pH_d	+ 4.606			0.251	0.551
	3	weight (%)	100					
		p value	0.003					
			-0.427 lg(K_SE)	+ 4.489			0.064	0.599
	4	weight (%)	100					
		p value	0.094					
			-0.936 lg(WC)	+ 4.212			0.016	0.642
	5	weight (%)	100					
		p value	0.236					
2-variable			0.405 lg(Cd_SE)	- 0.287 lg(Fe_SE)	+ 2.933		0.582	0.458
	1	weight (%)	87	13				
		p value	<0.001	0.005				
			0.389 lg(Cd_SE)	+ 0.424 lg(Si_SE)	+ 0.631		0.512	0.473
	2	weight (%)	94	6				
		p value	<0.001	0.111				
			0.392 lg(Cd_SE)	- 0.124 lg(Mn_SE)	+ 2.588		0.479	0.482
	3	weight (%)	97	3				
		p value	<0.001	0.249				
			0.349 lg(Cd_SE)	- 0.222 lg(K_SE)	+ 3.112		0.472	0.489
	4	weight (%)	89	11				
		p value	<0.001	0.262				
			0.313 lg(Cd_SE)	- 0.135 pH_m	+ 2.961		0.465	0.523
	5	weight (%)	66	34				
		p value	0.003	0.321				
3-variable			0.320 lg(Cd_SE)	- 0.323 lg(Fe_SE)	-0.211 pH_m	+ 4.293	0.603	0.452
	1	weight (%)	54	16	29			
		p value	<0.001	0.004	0.083			
			0.391 lg(Cd_SE)	- 0.300 lg(Fe_SE)	- 0.241 lg(K_SE)	+ 4.004	0.600	0.446
	2	weight (%)	77	13	10			
		p value	<0.001	0.004	0.143			
			0.401 lg(Cd_SE)	- 0.268 lg(Fe_SE)	+ 0.287 lg(Si_SE)	+ 1.853	0.588	0.462
	3	weight (%)	86	11	3			
		p value	<0.001	0.010	0.225			
			0.344 lg(Cd_SE)	- 0.327 lg(Fe_SE)	-0.147 pH_d	+ 3.906	0.578	0.478
	4	weight (%)	58	17	26			
		p value	<0.001	0.005	0.217			
			0.405 lg(Cd_SE)	- 0.301 lg(Fe_SE)	- 0.295 lg(clay)	+ 3.396	0.573	0.473
	5	weight (%)	85.8	13.8	0.4			
		p value	<0.001	0.005	0.479			

Table S9. Top five k-variable models for grain Cd built from combined soil parameter set.

			Grain Cd models: Combined parameter set					
			lg(Cd_RT) =				**R^2_{adj}**	**CV-RMSE**
1-variable			0.370 lg(Cd_SE)	+ 2.149			0.469	0.487
	1	weight (%)	100					
		p value	<0.001					
			0.472 lg(Cd_DSE)	+ 1.793			0.379	0.504
	2	weight (%)	100					
		p value	<0.001					
			0.115 lg(Cd_ST)	+ 2.681			-0.025	0.647
	3	weight (%)	100					
		p value	0.622					
			-0.430 pH_m	+ 5.006			0.299	0.547
	4	weight (%)	100					
		p value	<0.001					
			0.489 lg(Co_DSE)	+ 1.871			0.278	0.556
	5	weight (%)	100					
		p value	0.002					
2-variable			0.405 lg(Cd_SE)	- 0.287 lg(Fe_SE)	+ 2.933		0.582	0.458
	1	weight (%)	87	13				
		p value	<0.001	0.005				
			0.600 lg(Cd_DSE)	- 0.422 lg(Fe_SE)	+ 2.770		0.581	0.408
	2	weight (%)	80	20				
		p value	<0.001	<0.001				
			0.360 lg(Cd_SE)	+ 0.427 lg(Mg_ST)	+ 0.603		0.512	0.473
	3	weight (%)	88	12				
		p value	<0.001	0.092				
			0.733 lg(Cd_DSE)	- 0.327 lg(Zn_DSE)	+ 2.349		0.479	0.464
	4	weight (%)	85	15				
		p value	<0.001	0.011				
			0.389 lg(Cd_SE)	+ 0.424 lg(Si_SE)	+ 0.631		0.512	0.473
	5	weight (%)	94	6				
		p value	<0.001	0.111				
3-variable			0.788 lg(Cd_DSE)	- 0.360 lg(Fe_SE)	- 0.255 lg(Zn_DSE)	+ 3.052	0.654	0.385
	1	weight (%)	71	18	11			
		p value	<0.001	<0.001	0.014			
			0.692 lg(Cd_DSE)	- 0.438 lg(Fe_SE)	- 0.454 lg(S_DSE)	+4.765	0.646	0.403
	2	weight (%)	76	20	5			
		p value	<0.001	<0.001	0.028			
			0.385 lg(Cd_SE)	- 0.319 lg(Fe_SE)	+ 0.556 lg(K_ST)	+ 0.742	0.640	0.433
	3	weight (%)	73	14	13			
		p value	<0.001	0.002	0.024			
			0.697 lg(Cd_DSE)	- 0.419 lg(Fe_SE)	- 0.369 lg(Zn_ST)	+ 3.410	0.637	0.388
	4	weight (%)	77	19	5			
		p value	<0.001	<0.001	0.026			
			0.774 lg(Cd_DSE)	- 0.385 lg(Fe_SE)	- 0.255 lg(Zn_SE)	+ 3.158	0.633	0.391
	5	weight (%)	71	19	10			
		p value	<0.001	<0.001	0.035			

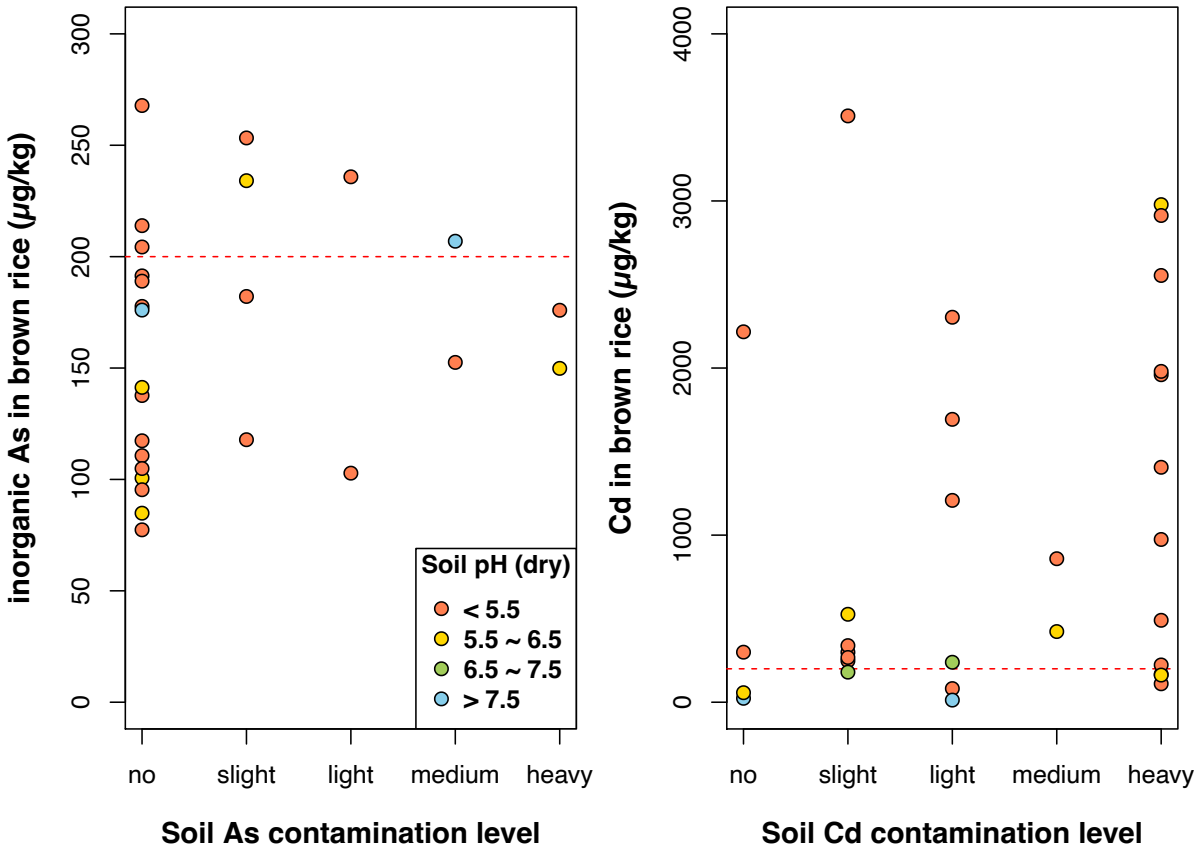

Figure S1. Inorganic As (n = 26) and total Cd (n = 31) in brown rice from rice paddies grouped into different soil contamination levels. Soil contamination levels are classified according to the Chinese national soil survey report in 2015 using risk screening values in soil environmental quality risk control standard for soil contamination of agricultural land (2018) as threshold value (T) for As and Cd, respectively: no (< T), slight (T – 2T), light (2T – 3T), medium (3T – 5T), heavy (> 5T). The threshold value (T) is set differently for different soil pH: For pH_d (dry soil pH) ⩽ 5.5, T is 0.3 mg/kg Cd and 30 mg/kg As in soil; For 5.5 < pH_d ⩽ 6.5, T is 0.4 mg/kg Cd and 30 mg/kg As in soil; For 6.5 < pH_d ⩽ 7.5, T is 0.6 mg/kg Cd and 25 mg/kg As in soil; For pH_d >7.5, T is 0.8 mg/kg Cd and 20 mg/kg As in soil. Horizontal red dash line stands for threshold values of Cd and iAs concentration in rice (both 200 µg/kg) for food safety in China. Data points are colored according to soil pH after drying.

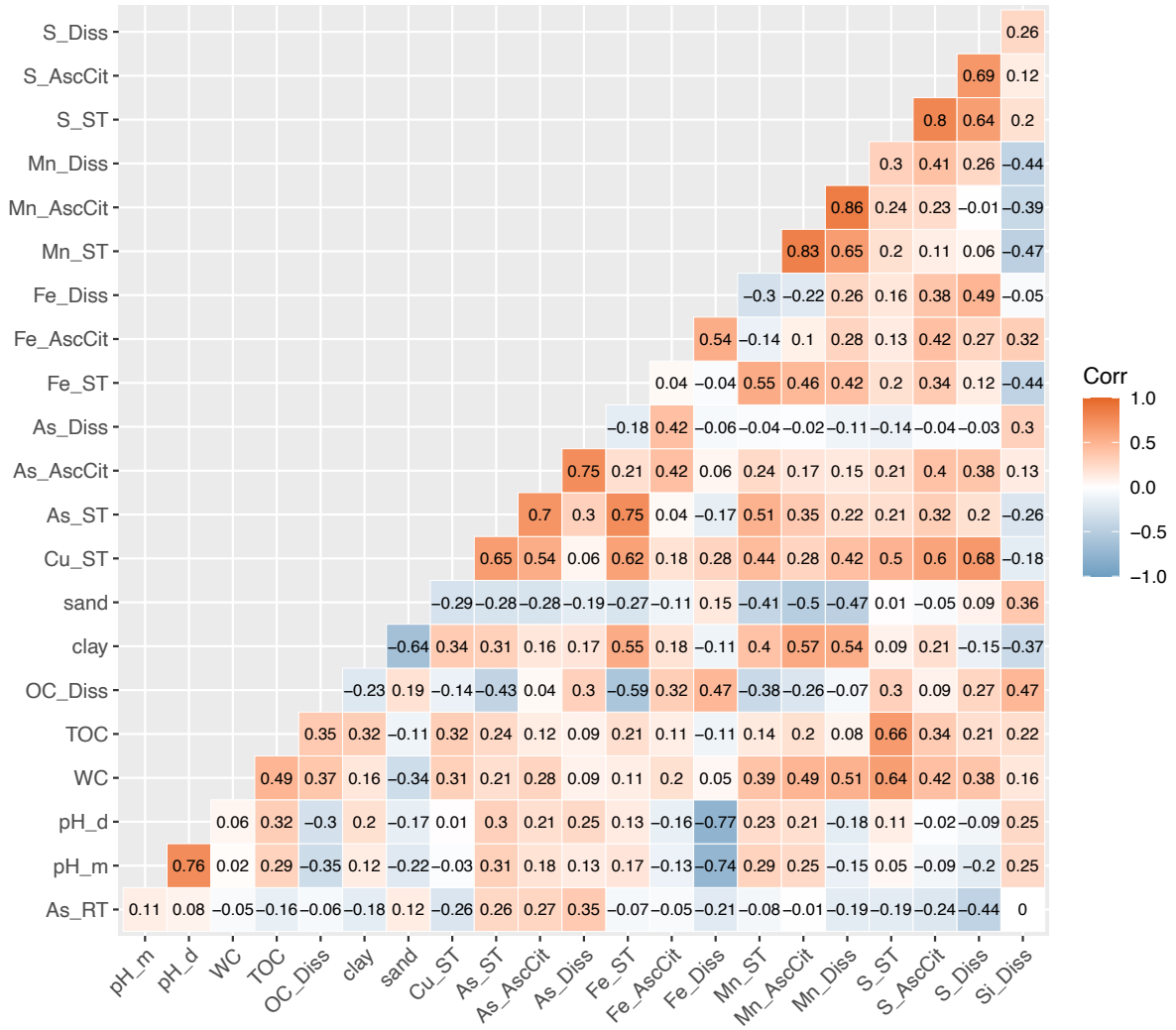

Figure S2. Correlation matrix of selected soil parameters in the variable pool for estimating grain As. Pearson's correlation coefficients are calculated for pairwise complete observations. Correlation to rice grain As from other variables (soil parameters) is in the bottom row. pH_m and pH_d: pH of field-moist soil and dry soil, respectively. WC: gravimetric water content. TOC: total organic carbon in soil. ST: soil total elemental concentration determined from aqua-regia digestion. AscCit: ascorbate-citrate-bicarbonate extractable concentration of dry soil. Diss: dissolved elemental concentration after 20-day anoxic soil incubation.

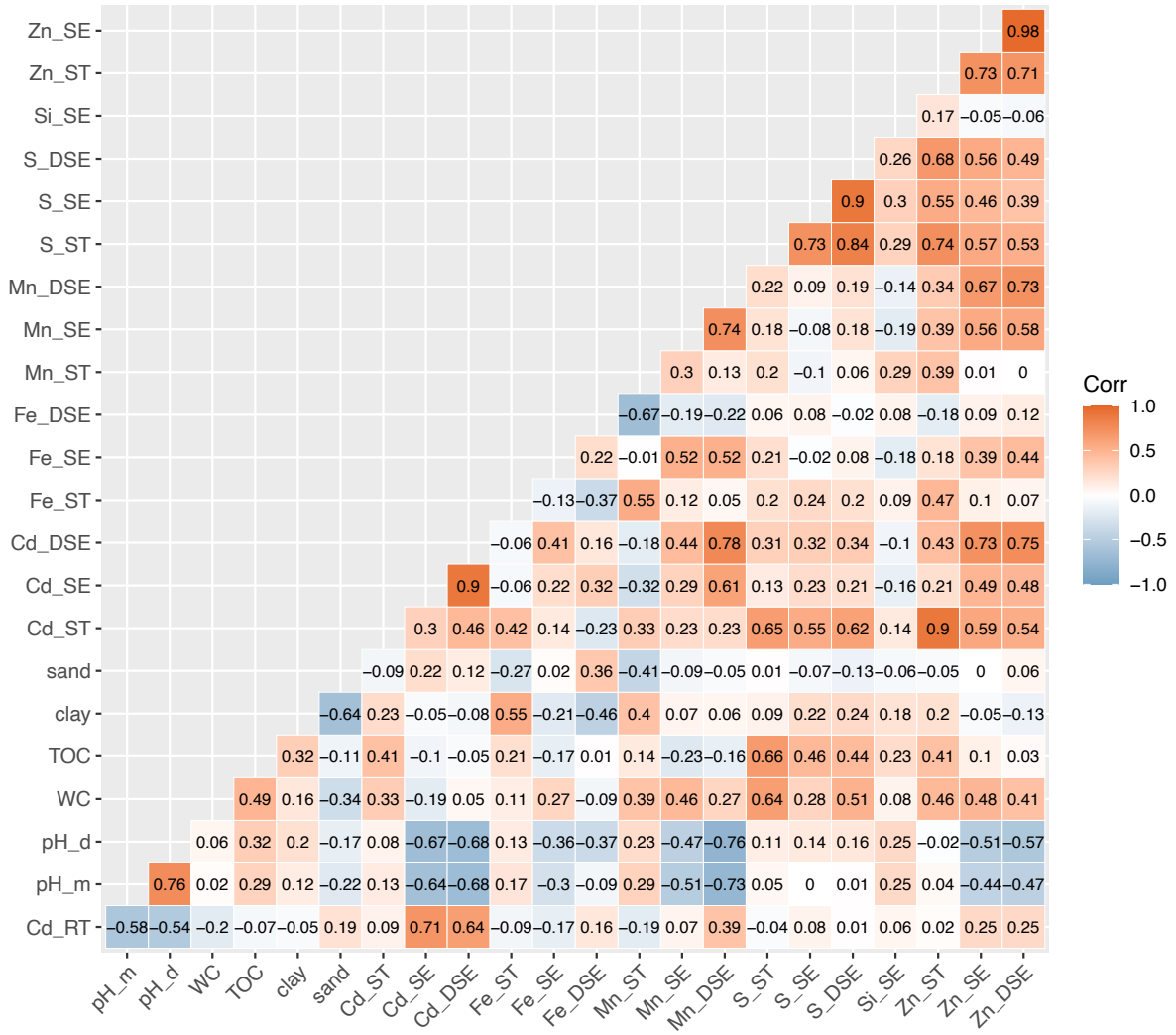

Figure S3. Correlation matrix of selected soil parameters in the variable pool for estimating grain Cd. Pearson's correlation coefficients are calculated for pairwise complete observations. Correlation to rice grain Cd from other variables (soil parameters) is in the bottom row. pH_m and pH_d: pH of field-moist soil and dry soil, respectively. WC: gravimetric water content. TOC: total organic carbon in soil. ST: soil total elemental concentration determined from aqua-regia digestion. SE: $CaCl_2$ extractable concentration of dry soil. DSE: $CaCl_2$ extractable concentration of field-moist soil at rice harvest.

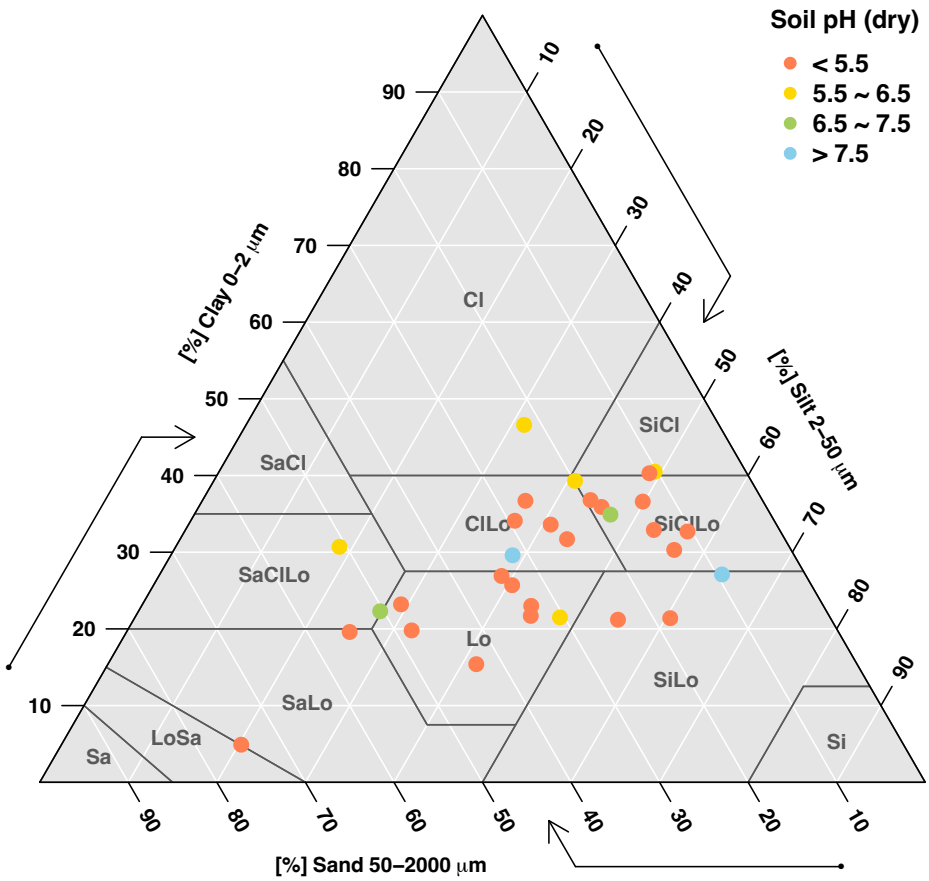

Figure S4. Classification for the 31 soils collected from this field survey. Soils were classified according to the USDA standards. Points were colored by the corresponding dry soil pH (0.01 M $CaCl_2$, 1:2.5 w/v). Cl: clay; SiCl: silty clay; SaCl: sandy clay; ClLo: clay loam; SiClLo: silty clay loam; SaClLo: sandy clay loam; Lo: loam; SiLo: silty loam; SaLo: sandy loam; Si: silt; LoSa: loamy sand; Sa: sand.