

Supplementary Material

Science of the Total Environment 2022, article no. 156058.
DOI: <https://doi.org/10.1016/j.scitotenv.2022.156058>

Monitoring and prediction of high fluoride concentrations in groundwater in Pakistan

Yuya Ling^a, Joel Podgorski^{a,*}, Muhammad Sadiq^b, Hifza Rasheed^c, Syed Ali Musstjab Akber Shah Eqani^b, Michael Berg^a

^a Eawag, Swiss Federal Institute of Aquatic Science and Technology, Department Water Resources and Drinking Water, 8600 Dübendorf, Switzerland

^b Public Health and Environment Division, Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan

^c National Water Quality Laboratory, Pakistan Council of Research in Water Resources (PCRWR), Islamabad, Pakistan

* Corresponding author: joel.podgorski@eawag.ch

CONTENTS

Figure S1	Box plot of fluoride concentrations measured in each province of Pakistan
Figure S2	Correlations of fluoride and other chemical parameters
Figure S3	Map of groundwater fluoride concentrations measured for this paper
Figure S4	Comparison of fluoride measurements in the field and at Eawag
Figure S5	Spatial distribution of the groundwater samples from different sources
Figure S6	Risk map of groundwater fluoride exceeding 1.5 mg/L
Figure S7	Risk maps of groundwater fluoride exceeding 1.0 mg/L
Figure S8	Combined fluoride prediction of Pakistan (this paper) and of the Indian Subcontinent
Figure S9	Variable importance plots under the fluoride threshold of 1.0 mg/L
Figure S10	ROC curve under the fluoride threshold of 1.0 mg/L
Figure S11	Maps of predictor variables
Table S1	Summary statistics of groundwater fluoride data used in modeling
Table S2	Summary statistics of groundwater fluoride concentrations measured for this paper
Table S3	Comparison of fluoride analyses using IC and the portable photometer field test kit
Table S4	Correlation and significance of the environmental predictors considered in modeling
Table S5	Distribution of fluoride levels in training and test sets under the threshold of 1.5 mg/L
Table S6	Distribution of fluoride levels in training and test sets under the threshold of 1.0 mg/L
Table S7	Pearson linear correlation coefficients between predictors

FIGURES

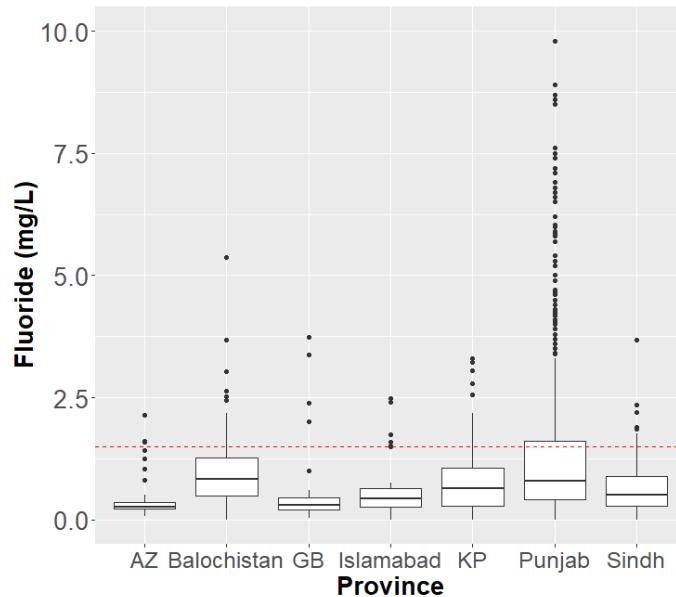


Figure S1. Box plots of fluoride concentrations measured in each province of Pakistan. The provinces are Azad Kashmir (AZ), Balochistan, Gilgit-Baltistan (GB), Islamabad, Khyber Pakhtunkhwa (KP), Punjab, and Sindh. The box plots show the medians (thick lines), quartiles (25, 75%) (box), whiskers (1.5 times the interquartile range), and the outliers. The red dashed line indicates the WHO guideline of 1.5 mg/L.

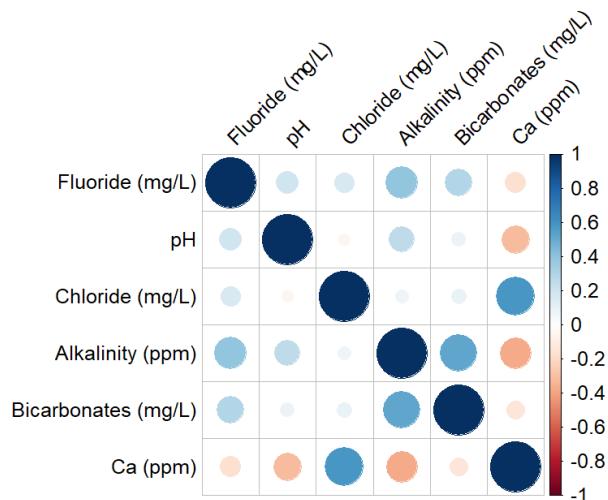


Figure S2. Correlation plot of fluoride, pH, chloride, alkalinity, bicarbonates, and calcium concentrations. The size of the dots correspond to the magnitude of correlation, with the coloring expressing positive (blue) and negative (red) values.

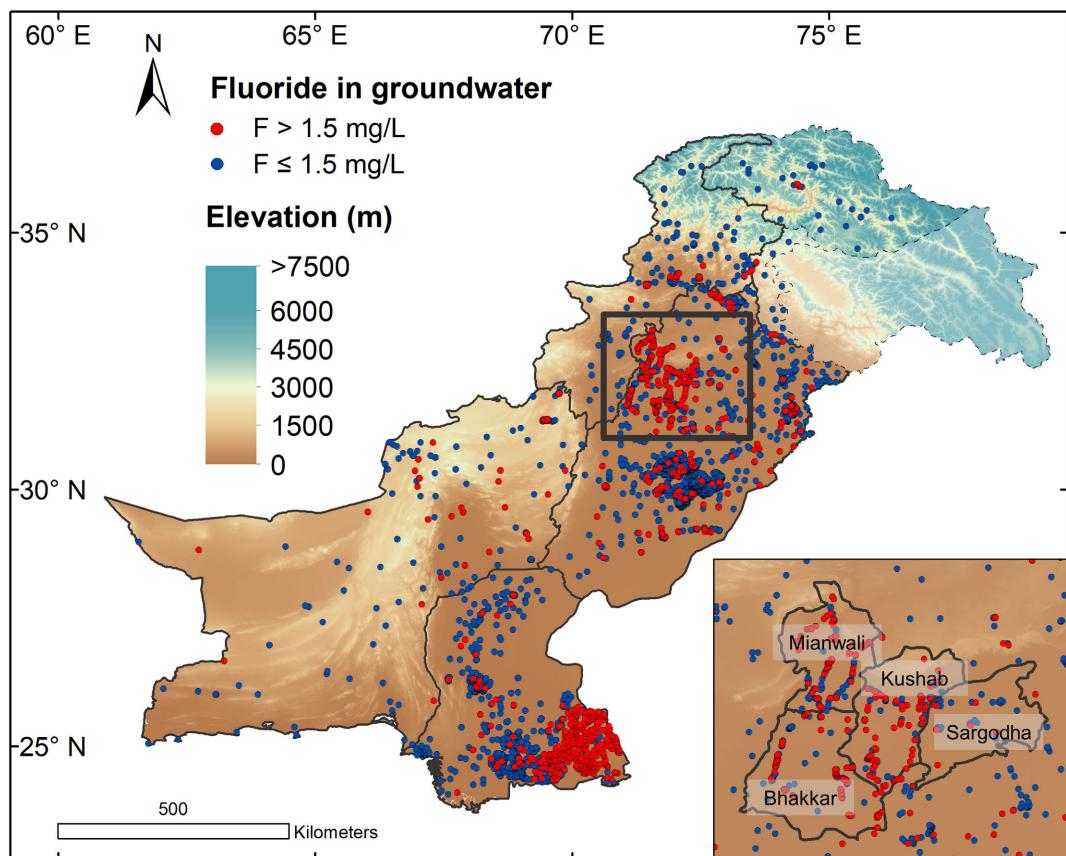


Figure S3. Topography of Pakistan plotted with groundwater fluoride concentrations ($n=2160$) measured for this paper by the co-authors at COMSATS Institute of Information Technology, Islamabad. Inset focuses on the Sargodha Division, of which the districts Bhakkar, Kushab and Mianwali are located in the upper Thal Desert.

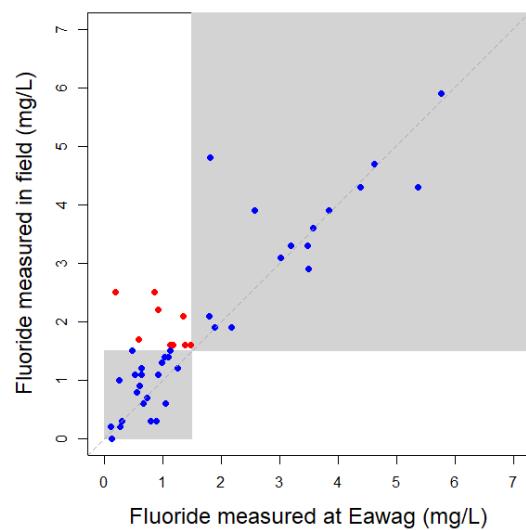


Figure S4. Comparison of fluoride measurements by FTK in the field and by IC in the laboratory. Shaded areas (blue dots, $n=39$) indicate fluoride concentrations measured by FTK and IC that are both either above or below 1.5 mg/L , whereas red dots ($n=9$) indicate inconsistently classified samples and were removed from the data set. However, four red dots are very close to the shaded areas (and being correctly classified).

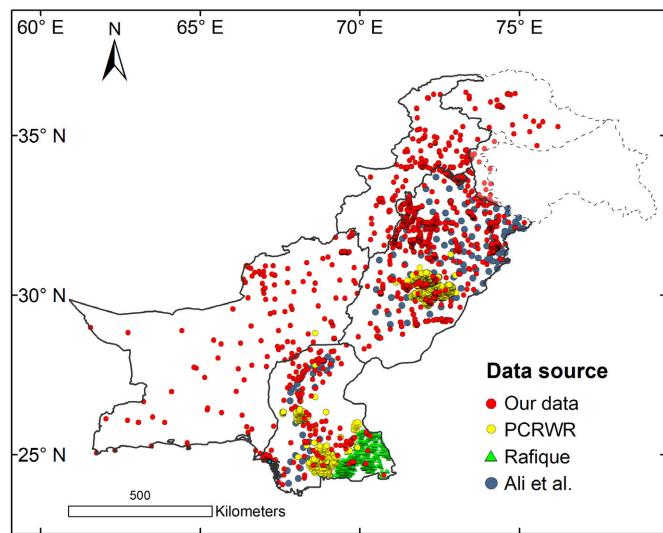


Figure S5. Spatial distribution of the groundwater samples from different sources.

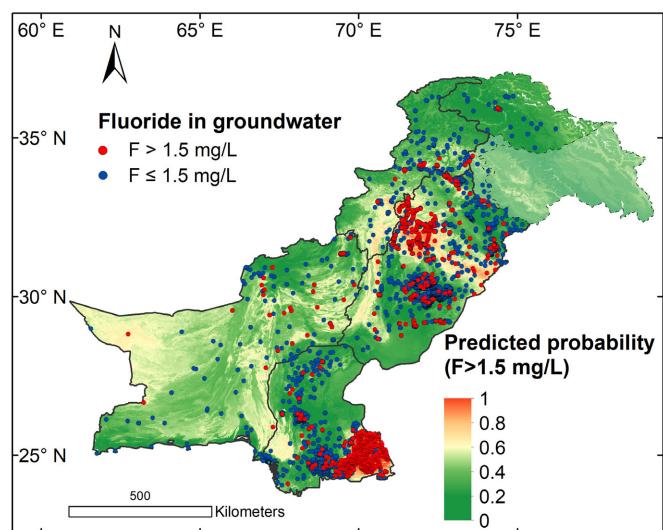


Figure S6. Risk map of groundwater fluoride exceeding 1.5 mg/L in Pakistan, shown along with measured fluoride concentrations.

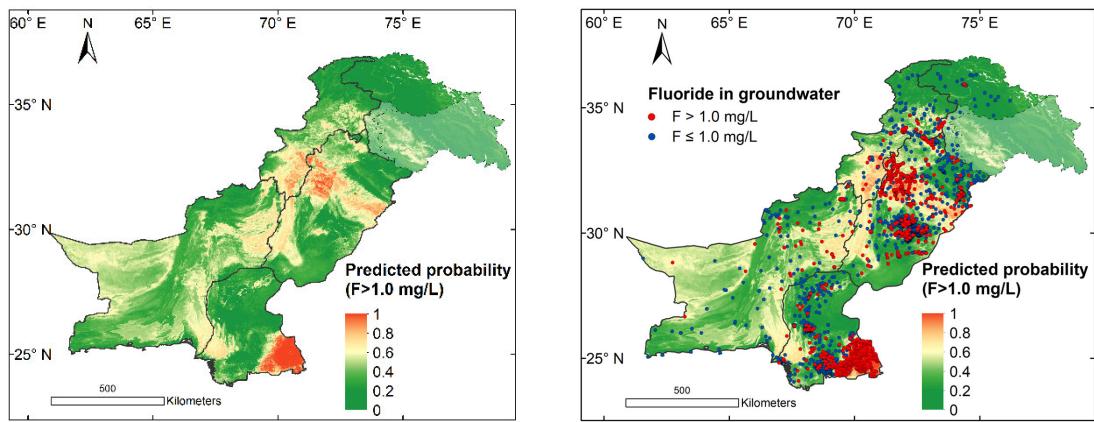


Figure S7. Risk maps of groundwater fluoride exceeding 1.0 mg/L in Pakistan, with and without measured fluoride concentrations.

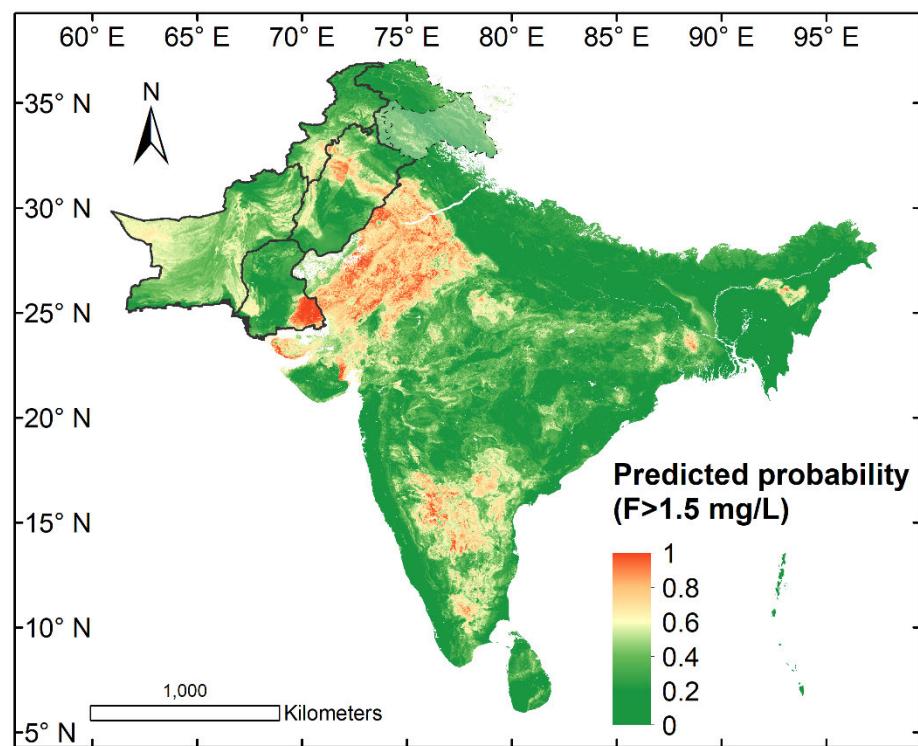


Figure S8. Fluoride prediction map of Pakistan (this paper) combined with a similar fluoride map of India and neighboring countries of South Asia (Podgorski, et al. 2018).

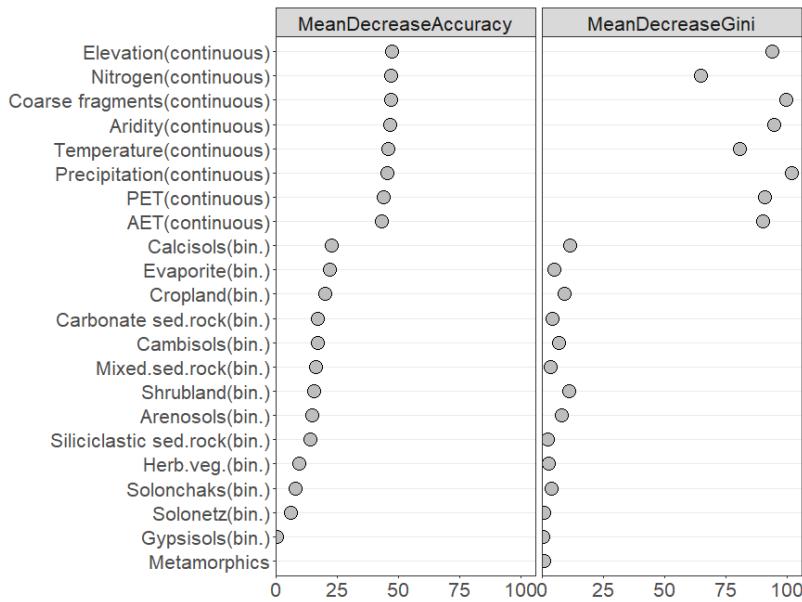


Figure S9. Variable importance plots of Mean Decrease Accuracy and Mean Decrease Gini Impurity for each variable in the random forest model under the threshold of 1.0 mg/L.

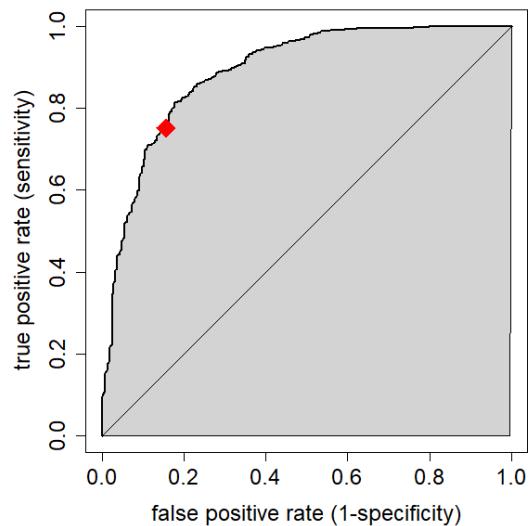


Figure S10. ROC curve of random forest model under the threshold of 1.0 mg/L (AUC score: 0.89).
The red diamond indicates the cut-off value of 0.43.

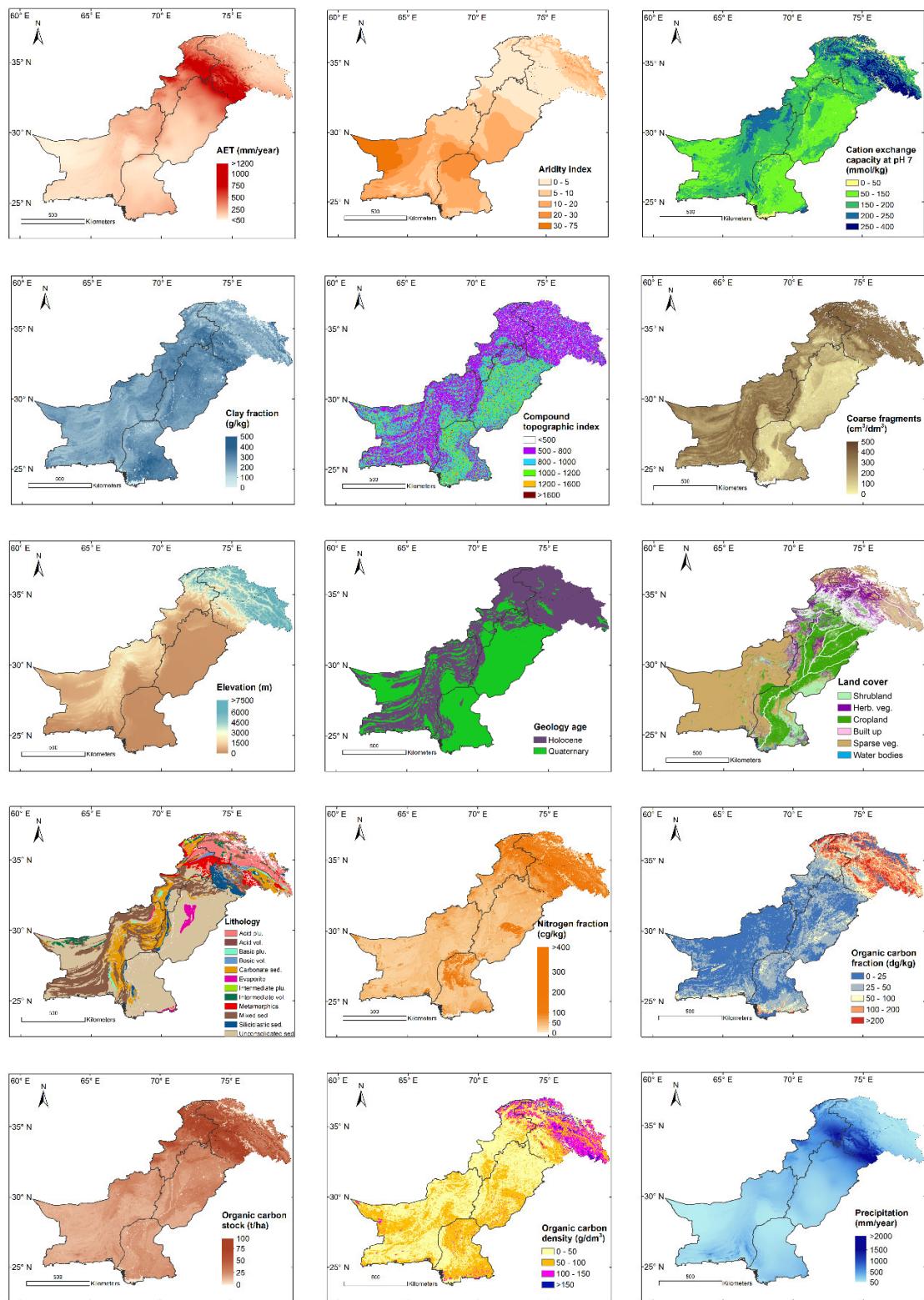


Figure S11. Maps of predictor variables considered in modeling (continued on next page).
Related information is provided in Table S4.

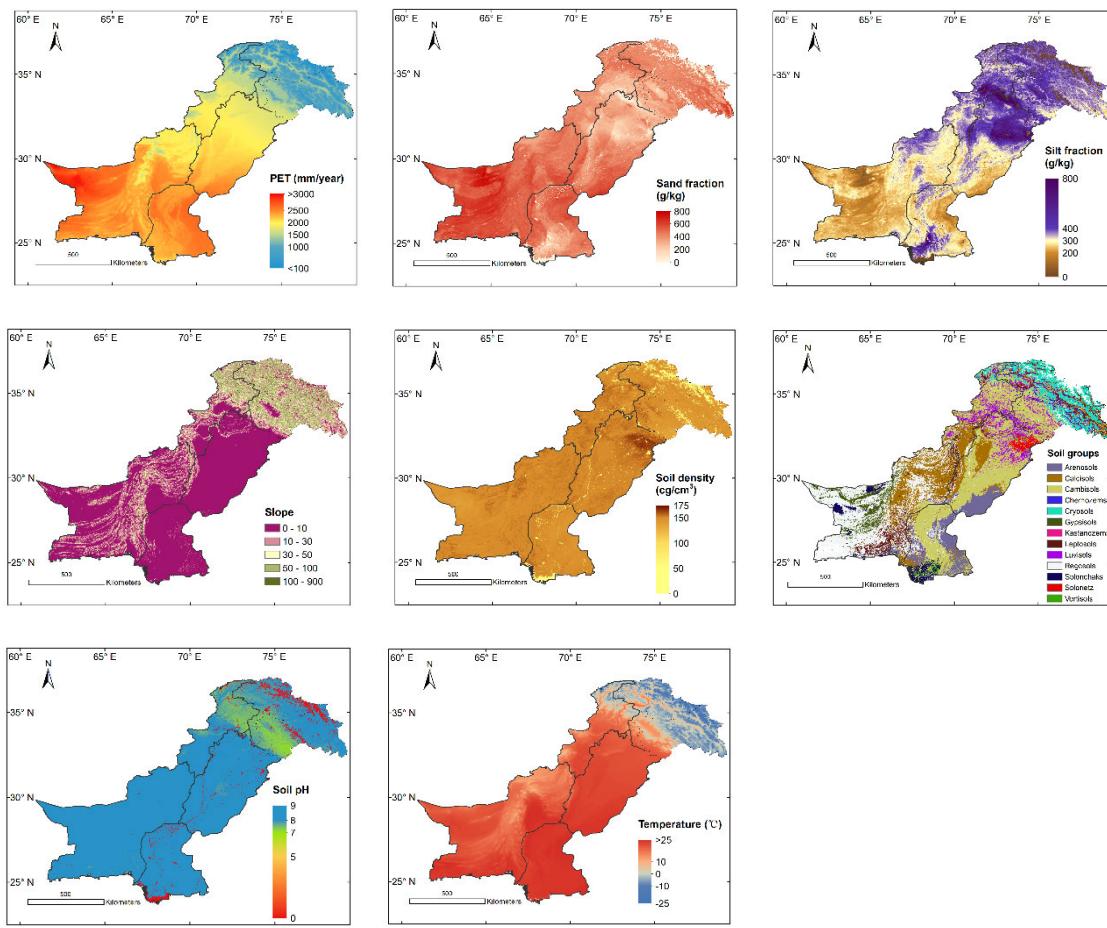


Figure S11 (cont.). Maps of predictor variables considered in modeling. Related information is provided in Table S4.

TABLES

Table S1. Summary statistics of groundwater fluoride data used in modeling.

Data source	No. of data	Max. / avg. ± std. F conc. (mg/L)	Study area
This paper	2160	33.3 / 1.11 ± 1.52	Samples from throughout Pakistan
Pakistan Council of Research in Water Resources (PCRWR 2015)	2814	7.50 / 0.54 ± 0.46	villages of Punjab and Sindh provinces
Ali, et al. (2019)	147	10.3 / 1.01 ± 1.24	along major rivers of Punjab and Sindh provinces
Rafique (2008)	422	30.25 / 3.78 ± 3.13	Thar desert

Table S2. Summary statistics of fluoride concentrations measured in the 2160 groundwater samples collected for this study.

Province	Division	District	F mean ± std (mg/L)	F median (mg/L)	F range (mg/L)	n
Azad Kashmir			0.47 ± 0.48	0.26	0.07 - 2.14	46
	Azad Kashmir		0.47 ± 0.48	0.26	0.07 - 2.14	46
		Bagh	0.25	0.25	0.25 - 0.25	1
		Bhimber	0.90 ± 0.80	0.49	0.18 - 2.14	7
		Kotli	0.14 ± 0.10	0.10	0.07 - 0.26	3
		Mirpur	0.26 ± 0.05	0.26	0.22 - 0.29	2
		Muzaffarabad	0.44 ± 0.40	0.28	0.21 - 1.59	30
		Neelum	0.32 ± 0.28	0.32	0.12 - 0.51	2
		Poonch	0.19	0.19	0.19 - 0.19	1
Baluchistan			0.93 ± 0.64	0.83	<0.1 - 5.36	245
	Kalat		0.76 ± 0.43	0.65	0.16 - 1.67	31
		Awaran	0.61 ± 0.10	0.64	0.50 - 0.70	3
		Kalat	0.48 ± 0.14	0.50	0.33 - 0.61	3
		Kharan	0.78 ± 0.21	0.74	0.58 - 1.17	6
		Khuzdar	0.76 ± 0.42	0.69	0.18 - 1.58	7
		Lasbela	0.58 ± 0.51	0.47	0.16 - 1.67	7
		Mastung	1.23 ± 0.51	1.56	0.65 - 1.65	5
	Makran		0.53 ± 0.34	0.46	0.18 - 1.64	16
		Gwadar	0.42 ± 0.05	0.45	0.36 - 0.47	5
		Kech	0.51 ± 0.21	0.47	0.29 - 0.86	8
		Panjgur	0.77 ± 0.77	0.49	0.18 - 1.64	3
	Nasirabad		0.90 ± 0.74	0.69	<0.1 - 2.44	16
		Bolan	1.10 ± 0.92	0.80	0.37 - 2.44	4
		Jafarabad	0.18 ± 0.10	0.21	<0.1 - 0.26	5
		Jhal Magsi	1.11 ± 0.51	0.99	0.57 - 1.81	5
		Nasirabad	1.78 ± 0.03	1.78	1.76 - 1.80	2
	Quetta		0.82 ± 1.10	0.31	<0.1 - 5.36	32
		Chagai	0.98 ± 0.81	0.66	<0.1 - 1.96	5
		Pishin	0.61 ± 0.68	0.28	<0.1 - 1.94	13

	Qilla Abdullah	0.17 ± 0.14	0.17	<0.1 - 0.41	8
	Quetta	1.98 ± 1.84	1.43	0.39 - 5.36	6
Sibi		1.03 ± 0.66	0.94	<0.1 - 3.68	40
	Dera Bugti	0.98 ± 0.63	0.91	0.30 - 3.68	30
	Kholu	1.26 ± 0.75	1.01	0.67 - 2.11	3
	Sibi	1.70 ± 0.48	1.73	1.18 - 2.17	4
	Ziarat	0.41 ± 0.53	0.23	<0.1 - 1.01	3
Zhob		1.04 ± 0.46	1.01	0.20 - 3.03	110
	Barkhan	1.33 ± 0.45	1.33	1.01 - 1.65	2
	Loralai	0.99 ± 0.58	0.83	0.26 - 1.60	5
	Musakhel	0.84 ± 0.50	0.64	0.50 - 1.58	4
	Qilla Saifullah	0.44 ± 0.20	0.44	0.20 - 0.66	4
	Zhob	1.07 ± 0.45	1.03	0.26 - 3.03	95
Gilgit-Baltistan		0.63 ± 0.93	0.30	0.03 - 3.73	31
Northern Areas		0.63 ± 0.93	0.30	0.03 - 3.73	31
	Chilas	0.25 ± 0.04	0.25	0.22 - 0.27	2
	Gilgit	0.88 ± 1.13	0.31	0.03 - 3.73	19
	Gilgit (Tribal Territory)	0.22	0.22	0.22 - 0.22	1
	Kargil	0.26 ± 0.11	0.22	0.14 - 0.44	7
	Kupwara (Gilgit Wazarat)	0.22	0.22	0.22 - 0.22	1
	Ladakh (Leh)	0.23	0.23	0.23 - 0.23	1
Islamabad		0.66 ± 0.66	0.43	<0.1 - 2.49	31
Islamabad		0.66 ± 0.66	0.43	<0.1 - 2.49	31
	Islamabad	0.66 ± 0.66	0.43	<0.1 - 2.49	31
Khyber Pakhtunkhwa		0.76 ± 0.63	0.63	<0.1 - 3.30	156
Bannu		0.96 ± 0.34	0.83	0.62 - 1.68	12
	Bannu	1.00 ± 0.37	0.87	0.62 - 1.68	9
	Lakki Marwat	0.83 ± 0.25	0.72	0.65 - 1.11	3
Dera Ismail Khan		0.69 ± 0.40	0.72	0.17 - 1.16	7
	Dera Ismail Khan	0.69 ± 0.43	0.73	0.17 - 1.16	6
	Tank	0.72	0.72	0.72 - 0.72	1
F.A.T.A.		0.51 ± 0.54	0.31	0.01 - 1.71	8
	Bhittani	0.64	0.64	0.64 - 0.64	1
	Khyber	0.36 ± 0.49	0.36	0.01 - 0.71	2
	Kurram	0.25	0.25	0.25 - 0.25	1
	Mohmand	0.25 ± 0.09	0.21	0.19 - 0.36	3
	Orakzai	1.71	1.71	1.71 - 1.71	1
Hazara	-	0.68 ± 0.67	0.39	0.02 - 2.78	27
	Abbottabad	1.12 ± 0.90	1.09	0.08 - 2.78	7
	Battagram	0.11	0.11	0.11 - 0.11	1
	Haripur	0.78 ± 0.63	0.64	0.10 - 1.89	6
	Kohistan	0.26 ± 0.12	0.23	0.15 - 0.47	5
	Mansehra	0.55 ± 0.51	0.38	0.02 - 1.38	8
Kohat		0.59 ± 0.39	0.45	0.13 - 1.25	11
	Hangu	0.13	0.13	0.13 - 0.13	1
	Karak	0.72 ± 0.44	0.96	0.21 - 1.25	7
	Kohat	0.45 ± 0.10	0.45	0.35 - 0.55	3
Malakand		0.48 ± 0.32	0.38	<0.1 - 1.05	29
	Chitral	0.40 ± 0.23	0.28	0.21 - 0.84	7

	Dir	0.35 ± 0.17	0.27	0.21 - 0.70	9
	Malakand P.A.	0.17	0.17	0.17 - 0.17	1
	Shangla	0.76	0.76	0.76 - 0.76	1
	Swat	0.65 ± 0.40	0.82	<0.1 - 1.05	11
Mardan		0.95 ± 0.84	0.72	<0.1 - 3.30	35
	Buner	0.84 ± 0.72	0.79	0.12 - 1.67	4
	Mardan	1.53 ± 0.87	1.25	0.65 - 3.23	7
	Swabi	0.80 ± 0.80	0.58	<0.1 - 3.30	24
Peshawar		0.99 ± 0.65	1.02	<0.1 - 2.56	27
	Charsadda	0.62 ± 0.77	0.23	<0.1 - 1.74	5
	Nowshera	1.02 ± 0.41	1.05	0.32 - 1.59	14
	Peshawar	1.18 ± 0.89	0.88	0.25 - 2.56	8
Punjab		1.18 ± 1.58	0.72	<0.1 - 27.51	1435
Bahawalpur		1.29 ± 1.89	0.70	<0.1 - 9.08	156
	Bahawalnagar	1.22 ± 1.90	0.46	<0.1 - 7.67	75
	Bahawalpur	1.40 ± 1.98	1.00	<0.1 - 9.08	74
	Rahimyar Khan	0.87 ± 0.54	0.99	0.17 - 1.78	7
Dera Ghazi Khan		0.47 ± 0.60	0.30	<0.1 - 2.63	122
	Dera Ghazi Kha	0.77 ± 0.67	0.47	0.16 - 2.63	45
	Layyah	0.26 ± 0.48	0.00	<0.1 - 2.00	42
	Muzaffargarh	0.30 ± 0.46	0.00	<0.1 - 1.50	32
	Rajan Pur	0.67 ± 0.50	0.47	0.30 - 1.23	3
Faisalabad		1.16 ± 1.70	0.77	<0.1 - 9.80	82
	Faisalabad	0.87 ± 0.50	0.74	0.17 - 2.01	19
	Jhang	1.29 ± 2.05	0.79	<0.1 - 9.80	54
	Toba Tek Singh	1.04 ± 0.65	0.88	0.43 - 2.41	9
Gujranwala		0.61 ± 0.30	0.61	<0.1 - 1.95	111
	Gujarat	0.61 ± 0.20	0.63	<0.1 - 0.86	26
	Gujranwala 1	0.68 ± 0.31	0.64	<0.1 - 1.95	43
	Gujranwala 2	0.28 ± 0.40	0.28	<0.1 - 0.56	2
	Gujrat	0.52 ± 0.38	0.49	<0.1 - 1.76	23
	Hafizabad	0.55 ± 0.19	0.49	0.38 - 0.86	5
	Narowal 1	0.5	0.50	0.50 - 0.50	1
	Narowal 2	0.92	0.92	0.92 - 0.92	1
	Sialkot	0.65 ± 0.32	0.59	0.19 - 1.40	10
Lahore		1.00 ± 2.25	0.60	<0.1 - 27.51	211
	Kasur	0.70 ± 0.67	0.69	<0.1 - 1.90	9
	Lahore	1.04 ± 1.51	0.68	0.08 - 12.80	102
	Nankana Sahib	0.98 ± 0.75	0.68	0.49 - 2.08	4
	Okara	0.64 ± 0.83	0.22	<0.1 - 1.80	6
	Okara 1	0.42 ± 0.92	0.00	<0.1 - 2.50	7
	Sheikhupura	1.06 ± 3.15	0.40	<0.1 - 27.51	83
Multan		0.74 ± 0.81	0.41	<0.1 - 4.25	189
	Khanewal	1.11 ± 0.63	1.02	0.09 - 2.95	44
	Lodhran	0.10 ± 0.17	0.01	<0.1 - 0.30	3
	Multan	0.46 ± 0.47	0.27	<0.1 - 1.91	41
	Pakpattan	0.05 ± 0.11	0.00	<0.1 - 0.30	8
	Sahiwal	0.61 ± 0.65	0.66	<0.1 - 1.70	11
	Vehari	0.79 ± 0.99	0.41	0.01 - 4.25	82
Rawalpindi		0.89 ± 0.79	0.56	<0.1 - 4.50	86
	Attok	0.26 ± 0.30	0.17	<0.1 - 0.85	6

	Chakwal	1.69 ± 1.26	1.26	0.17 - 4.50	12
	Jhelum	0.48 ± 0.29	0.42	<0.1 - 1.59	22
	Rawalpindi	0.96 ± 0.67	0.84	<0.1 - 2.36	46
Sargodha		1.76 ± 1.56	1.40	<0.1 - 10.50	478
	Bhakkar	1.54 ± 1.24	1.30	<0.1 - 7.20	203
	Khushab	2.50 ± 2.12	1.90	<0.1 - 10.50	121
	Mianwali	1.61 ± 1.18	1.50	<0.1 - 7.50	124
	Sargodha	0.94 ± 1.29	0.41	<0.1 - 5.87	30
Sindh		0.80 ± 2.28	0.54	<0.1 - 33.33	216
Hyderabad		0.44 ± 0.36	0.33	0.04 - 1.44	35
	Badin	0.17 ± 0.15	0.14	0.04 - 0.37	4
	Dadu	0.66 ± 0.46	0.56	0.22 - 1.31	4
	Hyderabad	0.24 ± 0.05	0.24	0.18 - 0.33	6
	Jamshoro	0.81 ± 0.59	0.80	0.17 - 1.44	6
	Matiari	0.39 ± 0.01	0.39	0.38 - 0.39	2
	Tando Allahyar	0.33 ± 0.24	0.33	0.16 - 0.50	2
	Tando M. Khan	0.71 ± 0.30	0.71	0.49 - 0.92	2
	Thatta	0.35 ± 0.07	0.33	0.25 - 0.47	9
Karachi		0.51 ± 0.39	0.34	0.07 - 1.43	35
	Karachi Central	0.46 ± 0.37	0.38	0.13 - 0.95	4
	Karachi East	0.29 ± 0.27	0.21	0.07 - 0.82	6
	Karachi South	0.53 ± 0.42	0.34	0.07 - 1.23	9
	Karachi west	0.39 ± 0.33	0.27	0.17 - 0.88	4
	Malir	0.65 ± 0.42	0.46	0.22 - 1.43	12
Larkana		0.66 ± 0.37	0.74	<0.1 - 1.71	45
	Jakobabad	0.48 ± 0.29	0.34	0.23 - 0.81	5
	Kashmore	0.54 ± 0.25	0.46	0.34 - 0.88	4
	Larkana	0.58 ± 0.33	0.72	<0.1 - 1.08	22
	Shikarpur	0.88 ± 0.43	0.86	0.11 - 1.71	14
Mirpur Khas		1.28 ± 3.83	0.76	0.03 - 33.33	74
	Mirphurkhas	0.60 ± 0.48	0.41	0.05 - 1.70	18
	Mithi	1.95 ± 5.59	0.89	0.03 - 33.33	34
	Sanghar	0.65 ± 0.49	0.50	0.11 - 1.60	12
	Umerkot	1.01 ± 0.65	0.85	0.18 - 2.35	10
Sukkur		0.56 ± 0.38	0.51	0.05 - 1.86	27
	Ghotki	0.35 ± 0.10	0.33	0.26 - 0.45	3
	Khairpur	0.44 ± 0.29	0.41	0.05 - 0.97	7
	Naushahro Firoz	0.80 ± 0.44	0.78	0.17 - 1.86	10
	Nawab Shah	0.44 ± 0.26	0.42	0.17 - 0.73	4
	Sukkur	0.42 ± 0.36	0.28	0.15 - 0.83	3

Table S3. Comparison of fluoride analyses using ion chromatography (IC) and the portable photometer field test kit (FTK) that was used to measure fluoride in the groundwater samples collected in this study. Five water samples were prepared with known concentrations of fluoride and subsequently measured by IC and the FTK.

Sample content	Sample preparation	Theoretical fluoride conc. (mg/L)	Fluoride conc. measured by IC (mg/L)	Fluoride conc. measured by FTK (mg/L)
Tap water (Eawag, Dübendorf)	Tap water	<0.1	0.07	0.05
ERM CA016a (certified F conc.: 1.5 ± 0.1 mg/L)	No dilution	1.5	1.6	1.57
IC Multi element standard; diluted 1:5 in tap water	10 mL standard solution has been topped up to 50 mL mark with tap water	2	2.1	2.13
F STD 1000 mg/L; diluted 1:125 in tap water	0.4 mg/L standard solution has been topped up to 50 mL mark with tap water	8	7.9	6.17
IC Multi element standard; diluted 1:20 in tap water	2.5 mg/L standard solution has been topped up to 50 mL mark with tap water	0.5	0.58	0.50

Table S4. Correlation and significance of the environmental predictors considered in modeling.

Type	Variable (Source)	Resolution	Depth	Correlation (p)
Soil (Continuous)	Bulk density (cg/cm ³) (Hengl, et al. 2017)	250 m	1 – 2 m	0.0166 (2.21E-01)
	Clay fraction (g/kg) (Hengl, et al. 2017)	250 m	1 – 2 m	0.0356 (8.57E-03)
	Sand fraction (g/kg) (Hengl, et al. 2017)	250 m	1 – 2 m	0.0401 (3.02E-03)
	Silt fraction (g/kg) (Hengl, et al. 2017)	250 m	1 – 2 m	-0.0319 (1.84E-2)
	Nitrogen fraction (cg/kg) (Hengl, et al. 2017)	250 m	1 – 2 m	-0.1363 (4.83E-24)
	Cation exchange capacity at pH 7 (mmol/kg) (Hengl, et al. 2017)	250 m	1 – 2 m	0.0053 (6.94E-01)
	Coarse fragments fraction (cm ³ /dm ³) (Hengl, et al. 2017)	250 m	1 – 2 m	0.2515 (1.75E-79)
	Organic carbon fraction (dg/kg) (Hengl, et al. 2017)	250 m	1 – 2 m	-0.0022 (8.68E-1)
	Organic carbon density (g/dm ³) (Hengl, et al. 2017)	250 m	1 – 2 m	0.1111 (1.91E-16)
	Organic carbon stock (t/ha) (Hengl, et al. 2017)	250 m	0 – 0.3 m	-0.1001 (1.22E-13)
Soil (Categorical)	Soil pH (Hengl, et al. 2017))	250 m	1 – 2 m	0.0435 (1.31E-3)
	Arenosols (Hengl, et al. 2017)	250 m	-	0.6677
	Calcisols (Hengl, et al. 2017)	250 m	-	0.3010
	Cambisols (Hengl, et al. 2017)	250 m	-	0.0993
	Chernozems (Hengl, et al. 2017)	250 m	-	0
	Cryosols (Hengl, et al. 2017)	250 m	-	0
	Gypsisols (Hengl, et al. 2017)	250 m	-	0
	Leptosols (Hengl, et al. 2017)	250 m	-	0.0769

	Luvisols (Hengl, et al. 2017)	250 m	-	0.1608
	Kastanozems (Hengl, et al. 2017)	250 m	-	0
	Regosols (Hengl, et al. 2017)	250 m	-	0.1006
	Solonchaks (Hengl, et al. 2017)	250 m	-	0.0376
	Solonetz (Hengl, et al. 2017)	250 m	-	0.0500
	Vertisols (Hengl, et al. 2017)	250 m	-	0.0652
Climate (Continuous)	Aridity index (Zomer, et al. 2007; Zomer, et al. 2008)	1000 m	surface	-0.1276 (3.09E-21)
	Actual evapotranspiration (AET) (mm/year) (Trabucco and Zomer 2010)	1000 m	surface	0.0276 (4.17E-02)
	Potential evapotranspiration (PET) (mm/year) (Zomer, et al. 2007; Zomer, et al. 2008)	1000 m	surface	0.1550 (1.07E-30)
	Precipitation (mm/year) (Fick and Hijmans 2017)	1000 m	surface	0.0412 (2.31E-03)
	Temperature (°C) (Fick and Hijmans 2017)	1000 m	surface	0.0902 (2.53E-11)
Geology-Age (Categorical)	Quaternary (Abu-Bakr, et al. 1964)	polygon	-	0.160
Lithology (Categorical)	Acid plutonics and volcanics (Hengl 2018)	250 m	-	0.1562
	Basic plutonics and volcanics (Hengl 2018)	250 m	-	0
	Carbonate sedimentary rocks (Hengl 2018)	250 m	-	0.2468
	Evaporite (Hengl 2018)	250 m	-	0.1644
	Intermediate plutionics and volcanics (Hengl 2018)	250 m	-	0
	Metamorphic rocks (Hengl 2018)	250 m	-	0.0566
	Mixed sedimentary rocks (Hengl 2018)	250 m	-	0.2025
	Siliciclastic sedimentary rocks (Hengl 2018)	250 m	-	0.0222
	Unconsolidated sediment (Hengl 2018)	250 m	-	0.1532
Topography (Continuous)	Elevation (m) (Verdin 2017)	100 m	surface	-0.0275 (4.22E-2)
	Slope (Verdin 2017)	100 m	surface	-0.0364 (7.12E-03)
	Compound topographic index (Verdin 2017)	100 m	surface	-0.0442 (1.10E-03)
Land Cover (Categorical)	Shrubland (Buchhorn, et al. 2020)	polygon	-	0.5664
	Built up (Buchhorn, et al. 2020)	polygon	-	0.1372
	Cropland (Buchhorn, et al. 2020)	polygon	-	0.0885
	Herbaceous vegetation (Buchhorn, et al. 2020)	polygon	-	0.3839
	Sparse/bare vegetation (Buchhorn, et al. 2020)	polygon	-	0.2007
	Water bodies (Buchhorn, et al. 2020)	polygon	-	0.0488

Table S5. Distribution of fluoride levels in training and test sets under the threshold of 1.5 mg/L.

	High	Low
Training set	703	3683
Test set	173	924

Table S6. Distribution of fluoride levels in training and test sets under the threshold of 1.0 mg/L.

	High	Low
Training set	1144	3242
Test set	275	822

Table S7. Pearson linear correlation coefficients between predictors.

	CEC	Clay	Coarse frag.	Density	Nitrogen	OC	OC density	OC stock	Soil pH	Sand	Silt	Aridity	PET	AET	Temp.	Precip.	Elevation	Slope	CTI
CEC	1	0.6397	0.4657	0.7728	0.1787	0.2447	0.4558	0.5468	0.7596	0.4730	0.5419	0.0625	0.0278	-0.1259	-0.2162	-0.0978	0.3162	0.1567	-0.0217
Clay	0.6397	1	0.1096	0.7708	0.1224	0.3800	0.5347	0.5012	0.7794	0.2305	0.5896	0.0333	0.2664	-0.2022	0.2106	-0.1607	-0.1635	-0.0313	0.0425
Coarse frag.	0.4657	0.1096	1	0.3125	-0.0095	0.1231	0.2715	0.2268	0.3022	0.4303	0.0375	-0.1320	0.0481	0.095	-0.3713	0.1151	0.5693	0.4574	-0.1836
Density	0.7728	0.7708	0.3125	1	0.2858	0.3638	0.5918	0.6094	0.9763	0.6540	0.6657	0.1838	0.2114	-0.2809	0.0613	-0.2361	-0.002	0.012	0.044
Nitrogen	0.1787	0.1224	-0.0095	0.2858	1	0.1709	0.3516	0.4769	0.2029	0.4057	-0.0056	0.1934	-0.0409	-0.19	-0.0335	-0.1779	-0.0228	-0.0093	0.0729
OC	0.2447	0.3800	0.1231	0.3638	0.1709	1	0.5277	0.4220	0.3317	0.233	0.2176	0.0967	0.1634	-0.1119	0.042	-0.0943	-0.0089	0.1438	0.0292
OC density	0.4558	0.5347	0.2715	0.5918	0.3516	0.5277	1	0.4118	0.5934	0.5455	0.2006	0.2823	0.4424	-0.3451	0.2488	-0.3191	-0.1241	-0.0127	0.1059
OC stock	0.5468	0.5012	0.2268	0.6094	0.4769	0.4220	0.4118	1	0.495	0.3228	0.4804	-0.111	-0.2492	0.1395	-0.3588	0.165	0.2955	0.3376	-0.0108
Soil pH	0.7596	0.7794	0.3022	0.9763	0.2029	0.3317	0.5934	0.495	1	0.625	0.6701	0.2541	0.3321	-0.3479	0.1616	-0.3062	-0.0637	-0.0376	0.0473
Sand	0.4730	0.2305	0.4303	0.6540	0.4057	0.2330	0.5455	0.3228	0.625	1	-0.0592	0.5132	0.3927	-0.5099	0.1189	-0.4866	0.0419	0.0074	0.053
Silt	0.5419	0.5896	0.0375	0.6657	-0.0056	0.2176	0.2006	0.4804	0.6701	-0.0592	1	-0.1905	-0.149	0.1387	-0.1204	0.1695	0.0416	0.0505	-0.0123
Aridity	0.0625	0.0333	-0.1320	0.1838	0.1934	0.0967	0.2823	-0.111	0.2541	0.5132	-0.1905	1	0.6887	-0.745	0.5117	-0.7672	-0.3378	-0.2094	0.1361
PET	0.0278	0.2664	0.0481	0.2114	-0.0409	0.1634	0.4424	-0.2492	0.3321	0.3927	-0.149	0.6887	1	-0.6366	0.7761	-0.6317	-0.5022	-0.2437	0.0797
AET	-0.1259	-0.2022	0.095	-0.2809	-0.19	-0.1119	-0.3451	0.1395	-0.3479	-0.5099	0.1387	-0.745	-0.6366	1	-0.5105	0.9851	0.3409	0.3317	-0.186
Temp.	-0.2162	0.2106	-0.3713	0.0613	-0.0335	0.042	0.2488	-0.3588	0.1616	0.1189	-0.1204	0.5117	0.7761	-0.5105	1	-0.5154	-0.9215	-0.5131	0.1445
Precip.	-0.0978	-0.1607	0.1151	-0.2361	-0.1779	-0.0943	-0.3191	0.165	-0.3062	-0.4866	0.1695	-0.7672	-0.6317	0.9851	-0.5154	1	0.3446	0.3234	-0.1815
Elevation	0.3162	-0.1635	0.5693	-0.002	-0.0228	-0.0089	-0.1241	0.2955	-0.0637	0.0419	0.0416	-0.3378	-0.5022	0.3409	-0.9215	0.3446	1	0.5695	-0.1717
Slope	0.1567	-0.0313	0.4574	0.012	-0.0093	0.1438	-0.0127	0.3376	-0.0376	0.0074	0.0505	-0.2094	-0.2437	0.3317	-0.5131	0.3234	0.5695	1	-0.322
CTI	-0.0217	0.0425	-0.1836	0.044	0.0729	0.0292	0.1059	-0.0108	0.0473	0.053	-0.0123	0.1361	0.0797	-0.186	0.1445	-0.1815	-0.1717	-0.322	1

References

- Abu-Bakr, A.M., Jackson, R.C. (1964). "Geological map of Pakistan, scale: 1:2,000,000." Ministry of Industries and Natural Resources, Geological Survey of Pakistan.
- Ali, W., et al. (2019). "Elucidating various geochemical mechanisms drive fluoride contamination in unconfined aquifers along the major rivers in Sindh and Punjab, Pakistan." Environmental Pollution **249**: 535-549.
- Buchhorn, M., et al. (2020). "Copernicus Global Land Service: Land Cover 100m: Collection 3 Epoch 2015, Globe(V3. 0.1) [Data set]." Zenodo. doi: 10.5281/zenodo.3939038
- Fick, S.E., Hijmans, R.J. (2017). "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas." International Journal of Climatology **37**(12): 4302-4315. doi: 10.1002/joc.5086
- Hengl, T. (2018). "Global landform and lithology class at 250 m based on the USGS global ecosystem map (1.0)." Zenodo. doi: 10.5281/zenodo.1464846.
- Hengl, T., et al. (2017). "SoilGrids250m: Global gridded soil information based on machine learning." Plos One **12**(2): e0169748. doi: 10.1371/journal.pone.0169748
- PCRWR (2015). Pakistan Council of Research in Water Resources (PCRWR), Islamabad, Pakistan.
- Podgorski, J., et al. (2018). "Prediction Modeling and Mapping of Groundwater Fluoride Contamination throughout India." Environmental Science & Technology **52**(17): 9889-9898. doi: 10.1021/acs.est.8b01679
- Rafique, T. (2008). Occurrence, distribution and origin of fluoride-rich groundwater in the Thar Desert, Pakistan. Sindh university jamshoro.
- Trabucco, A., Zomer, R.J. (2010). "Global soil water balance geospatial database." CGIAR Consortium for Spatial Information. <https://cgiarcsi.community/data/global-high-resolution-soil-water-balance/>
- Verdin, K. (2017). "Hydrologic derivatives for modeling and applications (HDMA) database: US Geological Survey data release." doi: 10.5066/F7S180ZP
- Zomer, R.J., et al. (2007). "Trees and water: smallholder agroforestry on irrigated lands in Northern India", International Water Management Institute, Colombo, Sri Lanka. IWMI Research Report 122, 47p.
- Zomer, R.J., et al. (2008). "Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation." Agriculture Ecosystems & Environment **126**(1-2): 67-80. doi: 10.1016/j.agee.2008.01.014