Supplementary materials

Controls on Groundwater Fluoride Contaminations in Eastern Parts of India: Insights from Unsaturated Zone Fluoride Profiles and AI-Based Modeling

David Anand Aind [1,†], Pragnaditya Malakar [1,2,†], Soumyajit Sarkar [3] and Abhijit Mukherjee [1,3,*]

[1] Department of Geology & Geophysics, Indian Institute of Technology Kharagpur, Kharagpur-721302, India
[2] Department of Geological Sciences, Jadavpur University, Kolkata-700032, India
[3] School of Environmental Sciences and Engineering, Indian Institute of Technology Kharagpur, Kharagpur-721302, India
* Correspondence: amukh2@gmail.com or abhijit@gg.iitkgp.ac.in
† These authors contributed equally to this work.

Text S1. Details of Predictors used for groundwater fluoride model

Climate variables: Higher fluoride levels have been corroborated to drier climate. Increased precipitation dilutes fluoride and lowers its concentration, but high temperatures and evapotranspiration combined with arid or semi-arid climate accelerate fluoride dissolution [3]. The climate factors were thus considered as potential indicators of groundwater fluoride. Climate factors have been identified as significant predictors of groundwater fluoride in earlier investigations [78]. The WorldClim database's "temperature" and "precipitation" climate variables were retrieved here with a spatial resolution of 30 arc-seconds [100]. The 2017 yearly average values of these variables were used. The Consultative Group for International Agriculture Research (CGIAR) database's "Potential Evapotranspiration" and "aridity" values were also retrieved with a 30 arc-second resolution [101].

Soil variables: In total, five soil factors were used as potential predictors in this study. These factors include the pH and cation exchange capacity of soil at two depths, namely the topsoil (at 50 cm depth) and subsoil (at 2m depth). Additionally, we used the percentage of fluvisols, a proxy of younger alluvial deposit in the soil (considered at a depth of 2 metres). The interaction of fluoride-minerals and groundwater can be influenced by soil variables like pH, and this can lead to fluoride dissolving into the groundwater [102]. By releasing or absorbing fluoride from their surfaces, soil variables can influence groundwater fluoride levels. Many previous studies have used soil variables as important predictors of groundwater fluoride [16, 103]. The International Soil Reference and Information Center (ISRIC)'s SoilGrids database was used to collect these variables at a resolution of 30 arc seconds [104].

Hydrogeologic variables: We have used four variables, i.e., “slope”, “topographic wetness index (TWI)”, “depth to groundwater”, and “hydraulic conductivity”, from the hydrogeologic category.

The variables "slope" and "topographic wetness index" are the proxies of groundwater flow rates, storage, and residence times and affect the interaction duration of fluoride-bearing minerals and groundwater [105]. The HydroSHEDS provides the global digital elevation model (DEM) database at 15 arc-second resolutions [106] and the DEM data from this database was used to develop the “slope” and “topographic wetness index” variables in the ArcGIS software at 30 arc-second resolutions.

"Depth to groundwater" can affect groundwater fluoride concentrations because changes in groundwater depth affect the pH and alkalinity of the water, which in turn encourages the dissolution of fluoride [107]. The Water Resources Information System (WRIS) database was used to get the

"depth to groundwater" data for the year 2017 [108]. Groundwater depth readings for the study area from the database was used here. The point data was interpolated using the inverse distance weighting approach, taking into account the 12 closest data points, to a resolution of 30 arc seconds ArcGIS software [109].

Fluoride enrichment is influenced by "hydraulic conductivity," which affects the rates of infiltration and recharging as well as the length of time that rocks and groundwater interact [110, 111]. From the study of Bhanja et al. [33], “hydraulic conductivity” data was retrieved as polygons and then extracted using the ArcGIS software.

Geologic variable: One of the most influencing variable for groundwater fluoride in this study was considered to be the "tectonics." The occurrence of fluoride-bearing minerals from deep earth and the processes involved in fluoride mobilisation are both governed by tectonic evolution [112-114]. The tectonic dataset was developed from the study of Parvez and Ram [115]; Jain et al. [116] using ArcGIS software.

Anthropogenic variable: The anthropogenic variable "area irrigated with groundwater" is a surrogate for extensive groundwater abstraction for irrigation, which depletes groundwater reserves and is likely to encourage fluoride-releasing mechanisms [76]. The information for "area irrigated with groundwater" was gathered using the Food and Agriculture Organization's (FAO) Global Map of Irrigation Areas (GMIA), version 5 database, which has a resolution of 5 arc-minutes [117].

Table S1. List of variables selected for possible influence on groundwater Fluoride, their sources and relation with Fluoride based on previous literature.

Category	Predictor variables	Resolution	Type
Climate	Precipitation (P) (mm/year)	30 arc-second	Continuous
	Temperature (ºc)	30 arc-second	Continuous
	Potential evapotranspiration (PET) (mm/year)	30 arc-second	Continuous
	Aridity (P/PET)	30 arc-second	Continuous
Soil	Topsoil pH (%)	30 arc-second	Continuous
	Subsoil pH (%)	30 arc-second	Continuous
	Topsoil Cation exchange capacity (CEC_topsoil) (%)	30 arc-second	Continuous
	Subsoil Cation exchange capacity (CEC_subsoil) (%)	30 arc-second	Continuous
	Fluvisols (%)	30 arc-second	Continuous
Hydrogeology	Slope (º)	30 arc-second	Continuous
	Topographic wetness index (TWI)	30 arc-second	Continuous
	Depth to groundwater (m)	30 arc-second	Continuous

	Hydraulic conductivity (m/day)	Polygon	Categorical
Geology	Tectonics	Polygon	Categorical
Anthropogenic	Area irrigated with groundwater (%)	5 arc-minute	Continuous

Table S2. Details of the Random forest model performance

Performance parameter	Dataset		
	Train	Test	Overall
Accuracy	91%	90%	93%
Sensitivity	0.90	0.88	0.93
Specificity	0.99	0.99	0.92
AUC		0.95	0.99
Cutoff		0.55	0.58

Table S3: Details of groundwater F measurements (*n*=205) from CGWB database and RF predicted area with elevated (>1.0 mg/L) groundwater F and estimated population exposed to elevated groundwater F for the 3 studied districts of WB

Districts (Geological terrain)	Observed no of measurements >1.0 mg/L (total measurements)	Predicted	
		% of area with >1.0 mg/L F	Population exposed to >1.0 mg/L F
Bankura (P. Alluvium)	23 (87)	4	170559
P. Medinipur (Residual Soils)	3 (57)	1.8	114383
Puruliya (CGC)	0 (61)	2.6	106882

Table S4. Elemental oxide weight % of soil sediments derived from XRF analysis.

Samples	**Na_2O**	**MgO**	**Al_2O_3**	**SiO_2**	**P_2O_5**	**K_2O**	**CaO**	**TiO_2**	**Fe_2O_3**	**F**	**Cl**	**Total wt.%**
Bankura	**0.68**	**0.19**	**8.50**	**43.22**	**0.41**	**2.32**	**0.65**	**0.91**	**7.07**	**35.94**	**0.11**	
BK1-3-3.5	0.29	0.17	5.98	49.43	0.56	1.29	0.56	0.82	3.90	36.93	0.09	100
BK1-3-12	1.04	0.33	13.02	35.50	0.25	7.34	0.89	1.09	5.49	34.94	0.14	100
BK1-3-14	1.69	0.16	8.28	41.19	0.44	1.31	0.52	0.88	13.96	31.57	0.02	100
BK-3-4.5	0.30	0.19	7.00	47.64	0.47	1.43	0.84	0.88	4.32	36.92	0.02	100
BK-3-18	0.34	0.16	7.97	44.13	0.44	1.26	0.53	0.80	5.93	38.43	0.02	100
BK-3-22	0.42	0.17	8.77	41.47	0.29	1.28	0.58	0.99	8.82	36.84	0.37	100
Puruliya	**0.90**	**0.35**	**12.89**	**38.47**	**0.38**	**8.16**	**0.60**	**0.87**	**5.78**	**31.29**	**0.29**	

P-1-10	1.12	0.27	12.04	41.50	0.45	9.32	0.05	0.65	5.94	28.36	0.30	100
P-1-20	0.83	0.37	13.87	37.84	0.34	7.85	0.83	1.05	5.73	30.74	0.56	100
P-1-22	0.75	0.41	12.77	36.08	0.34	7.32	0.93	0.92	5.68	34.78	0.03	100
P. Medinipur	**0.51**	**0.63**	**9.65**	**41.94**	**0.40**	**1.60**	**3.38**	**1.27**	**7.75**	**32.70**	**0.17**	
WM3-1-4.5	0.51	0.45	7.83	47.87	0.43	1.23	0.98	1.14	6.08	33.28	0.22	100
WM3-1-12	0.42	0.69	10.76	40.92	0.48	1.74	1.20	1.31	8.45	33.75	0.29	100
WM3-1-16	0.60	0.91	9.78	42.29	0.48	2.03	4.05	1.20	8.66	29.96	0.05	100
WM3-2-8	0.42	0.59	9.49	39.42	0.29	1.55	5.90	1.32	7.74	33.21	0.09	100
WM3-2-12	0.39	0.45	10.83	41.38	0.32	1.40	0.97	1.36	7.02	35.72	0.17	100
WM3-2-20	0.71	0.72	9.21	39.77	0.40	1.63	7.18	1.27	8.57	30.32	0.23	100

	Area irrigated with groundwater	Precipitation	Temperature	Aridity	PET	Slope	TWI	Depth to groundwater	Fluvisols	Subsoil pH	Topsoil pH	CEC_subsoil	CEC_topsoil
Area irrigated with groundwater	1.00	0.05	-0.14	-0.02	0.15	0.08	0.00	0.02	-0.42	-0.30	-0.30	-0.20	-0.27
Precipitation	0.05	1.00	-0.36	0.78	-0.31	-0.05	0.26	0.03	0.37	0.40	0.22	-0.15	-0.09
Temperature	-0.14	-0.36	1.00	0.11	-0.56	-0.25	0.05	-0.08	0.25	0.29	0.17	-0.23	-0.13
Aridity	-0.02	0.78	0.11	1.00	-0.77	-0.14	0.25	0.00	0.57	0.55	0.33	-0.32	-0.19
PET	0.15	-0.31	-0.56	-0.77	1.00	0.25	-0.17	0.00	-0.67	-0.59	-0.42	0.40	0.23
Slope	0.08	-0.05	-0.25	-0.14	0.25	1.00	-0.33	0.02	-0.26	-0.28	-0.16	0.15	0.07
TWI	0.00	0.26	0.05	0.25	-0.17	-0.33	1.00	0.01	0.31	0.33	0.21	-0.03	0.02
Depth to groundwater	0.02	0.03	-0.08	0.00	0.00	0.02	0.01	1.00	0.10	0.09	0.00	0.01	0.03
Fluvisols	-0.42	0.37	0.25	0.57	-0.67	-0.26	0.31	0.10	1.00	0.72	0.64	-0.08	0.10
Subsoil pH	-0.30	0.40	0.29	0.55	-0.59	-0.28	0.33	0.09	0.72	1.00	0.68	-0.02	0.14
Topsoil pH	-0.30	0.22	0.17	0.33	-0.42	-0.16	0.21	0.00	0.64	0.68	1.00	0.07	0.19
CEC_subsoil	-0.20	-0.15	-0.23	-0.32	0.40	0.15	-0.03	0.01	-0.08	-0.02	0.07	1.00	0.95
CEC_topsoil	-0.27	-0.09	-0.13	-0.19	0.23	0.07	0.02	0.03	0.10	0.14	0.19	0.95	1.00

Figure S1. Correlation among the continuous predictor variables used in the Random forest model

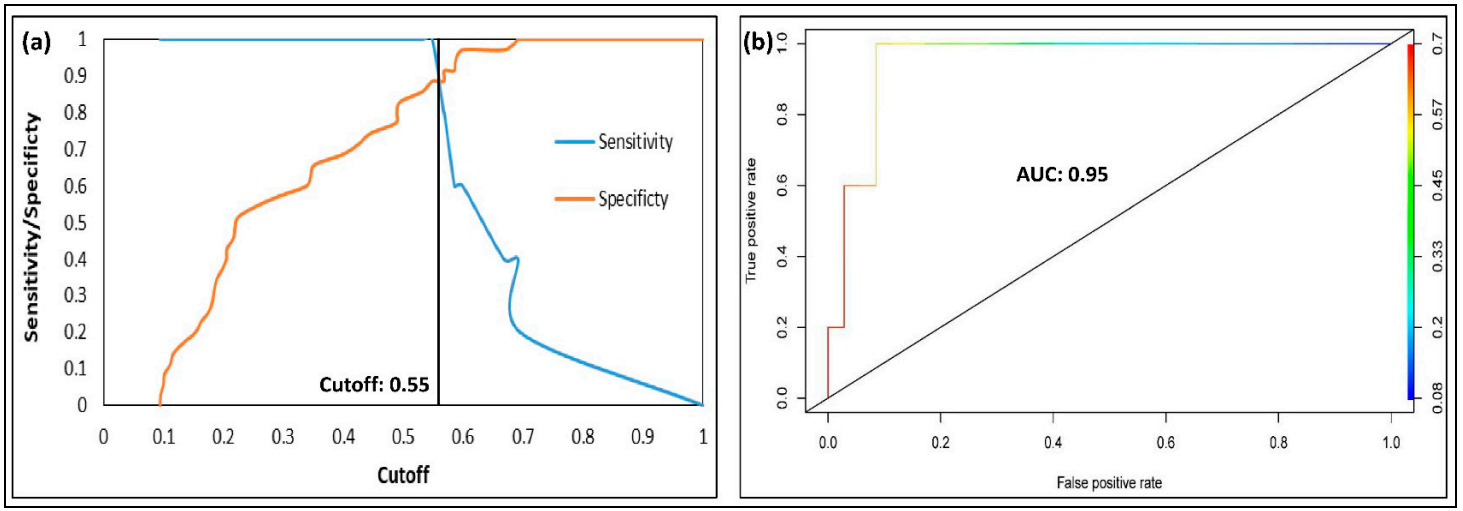

Figure S2. (a) Sensitivity vs. specificity of the Random forest model and selected cutoff for the test data, (b) Area under the curve (AUC) value determined from the Receiver operating characteristic (ROC) curve of the Random forest model for the test data

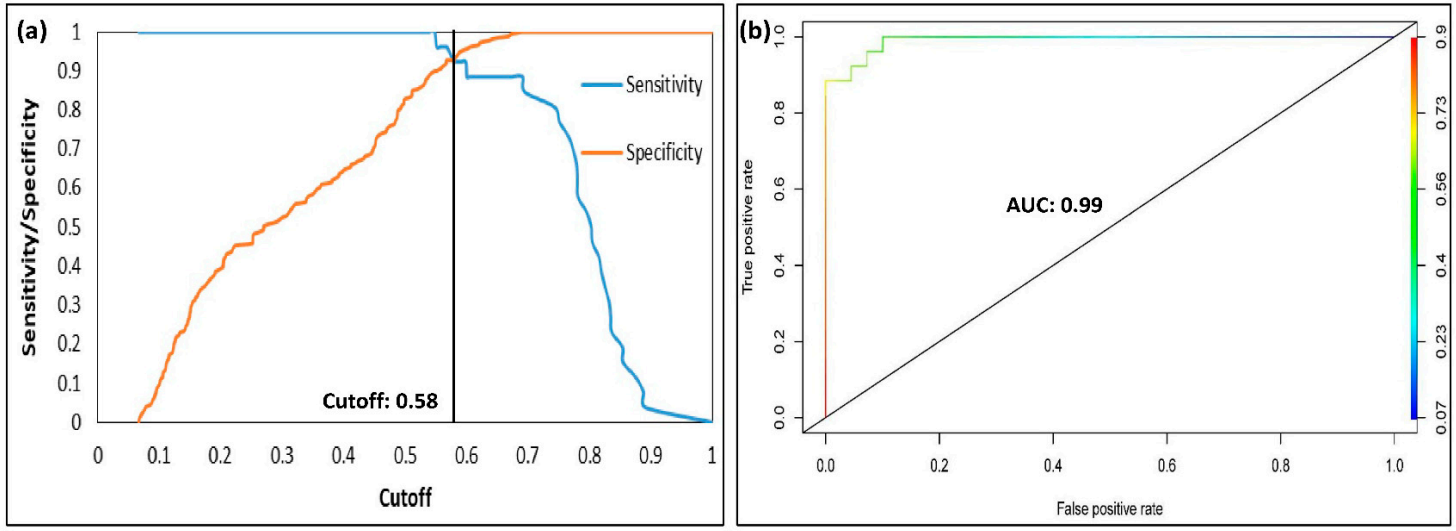

Figure S3. (a) Sensitivity vs. specificity of the Random forest model and selected cutoff for the overall data, (b) Area under the curve (AUC) value determined from the Receiver operating characteristic (ROC) curve of the Random forest model for the entire data

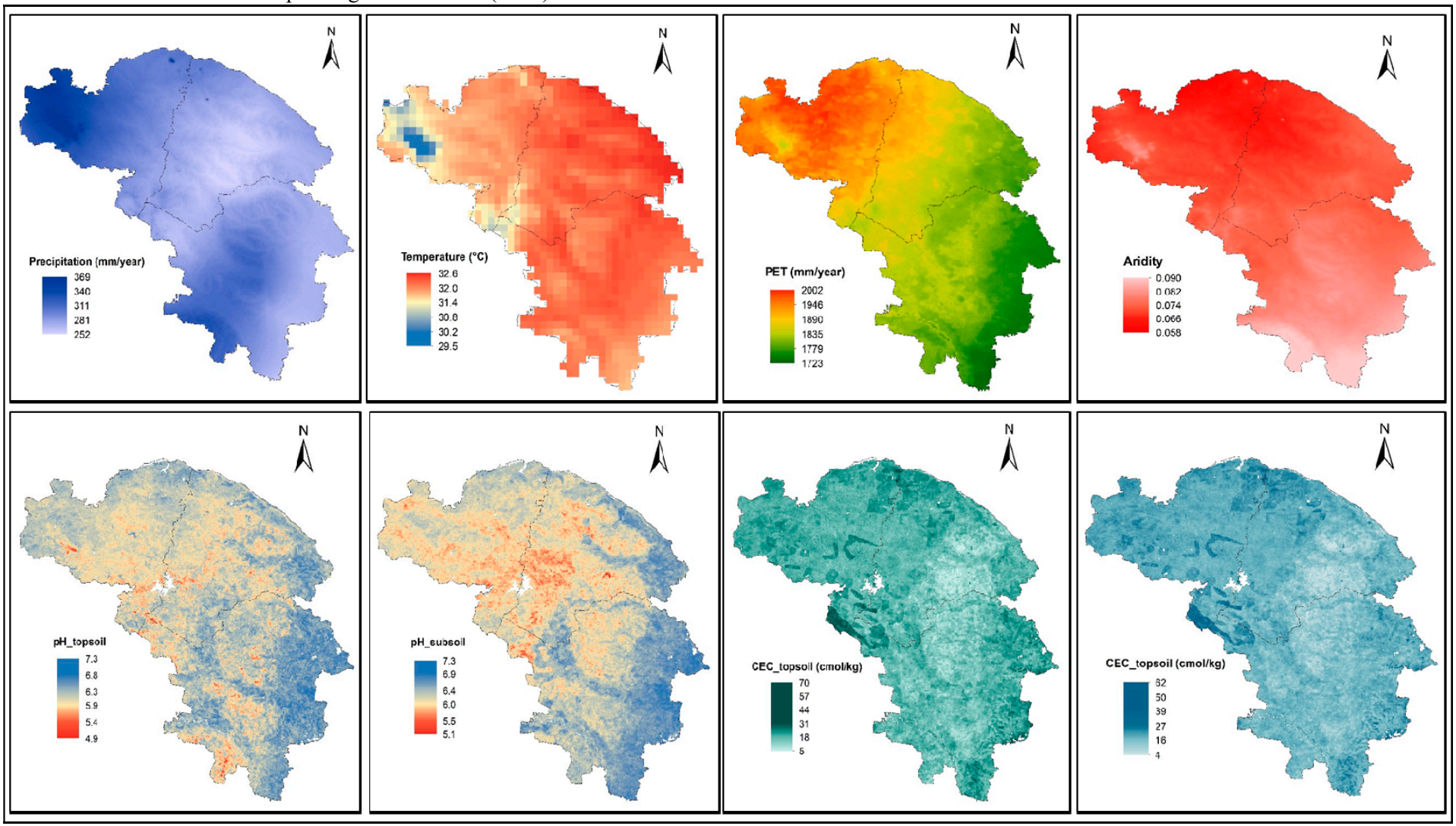

Figure S4. Maps of the predictor variables used in the Random forest model

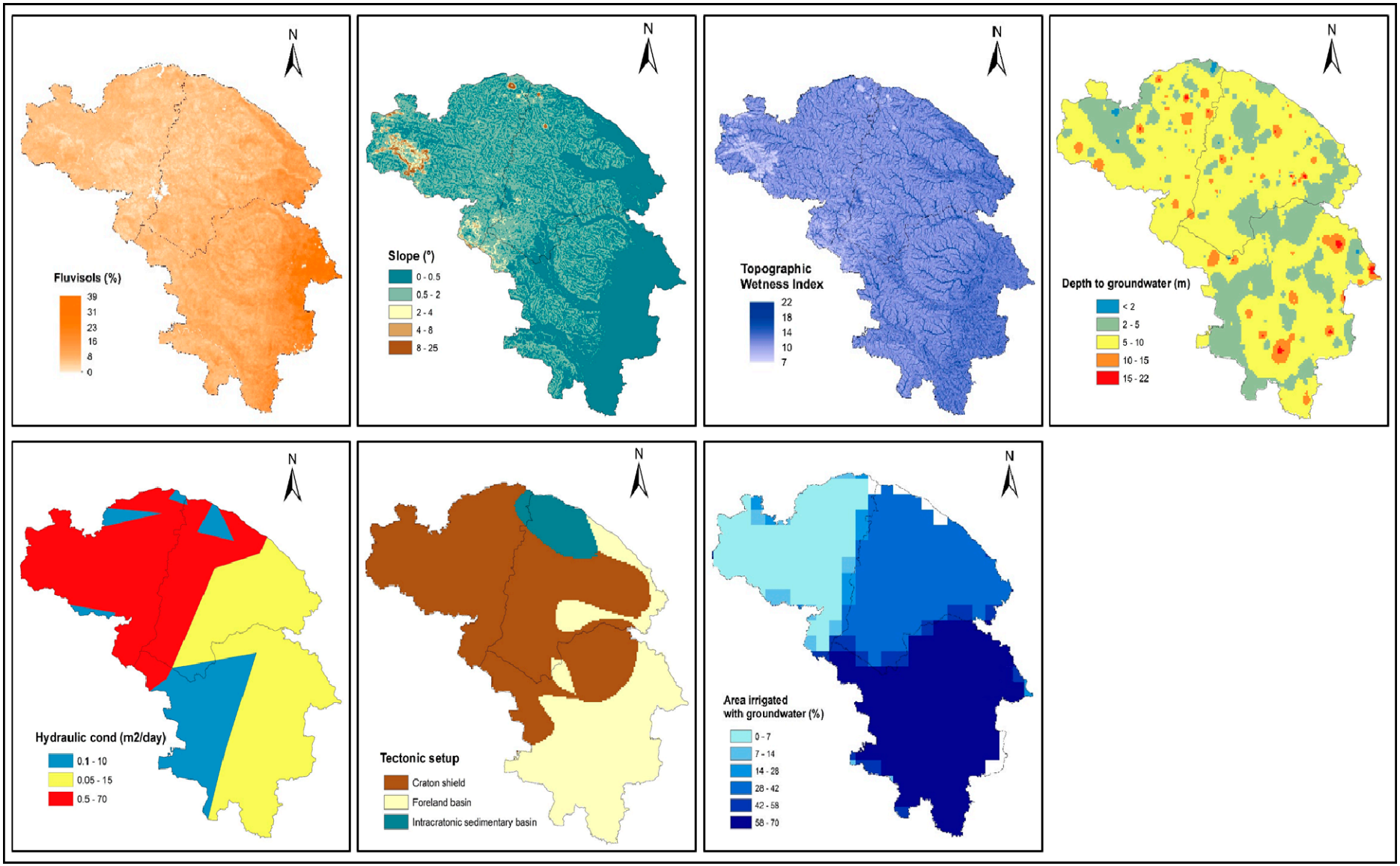

Figure S4. Maps of the predictor variables used in the Random forest model (Continued)

References

3. Edmunds, W.M.; Smedley, P.L. Fluoride in Natural Waters. In Essentials of Medical Geology: Revised Edition; Springer The Netherlands: Dordrecht, 2013; pp. 311–336 ISBN 9789400743755.

16. Podgorski, J.E.; Labhasetwar, P.; Saha, D.; Berg, M. Prediction Modeling and Mapping of Groundwater Fluoride Contamination throughout India. *Environ. Sci. Technol.* **2018**, *52*, 9889–9898, doi:10.1021/acs.est.8b01679

33. Bhanja, S.N.; Mukherjee, A.; Rangarajan, R.; Scanlon, B.R.; Malakar, P.; Verma, S. Long-Term Groundwater Recharge Rates across India by in Situ Measurements. Hydrol. Earth Syst. Sci. 2019, 23, 711–722, doi:10.5194/hess-23-711-2019.

76. Liu, Y.; Jin, M.; Ma, B.; Wang, J. Distribution and Migration Mechanism of Fluoride in Groundwater in the Manas River Basin, Northwest China. Hydrogeol. J. 2018, 26, 1527–1546, doi:10.1007/s10040-018-1780-8.

78. Cao, H.; Xie, X.; Wang, Y.; Liu, H. Predicting Geogenic Groundwater Fluoride Contamination throughout China. J. Environ. Sci. (China) 2022, 115, 140–148, doi:10.1016/j.jes.2021.07.005.

100. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315, doi:10.1002/joc.5086.

101. Trabucco, A. Zomer, R. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database V2 Available online: https://doi.org/10.6084/m9.figshare.7504448.v3 (accessed on 12 August 2022).

102. Sivasankar, V.; Darchen, A.; Omine, K.; Sakthivel, R. Fluoride: A World Ubiquitous Compound, Its Chemistry, and Ways of Contamination. In *Surface Modified Carbons as Scavengers for Fluoride from Water*; Springer International Publishing: Cham, 2016; pp. 5–32 ISBN 9783319406862.

103. Thapa, R.; Gupta, S.; Gupta, A.; Reddy, D.V.; Kaur, H. Geochemical and Geostatistical Appraisal of Fluoride Contamination: An Insight into the Quaternary Aquifer. *Sci. Total Environ.* **2018**, *640–641*, 406–418, doi:10.1016/j.scitotenv.2018.05.360.

104. Hengl, T.; De Jesus, J.M.; Heuvelink, G.B.M.M.; Gonzalez, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. **2017**, *12*, e0169748, doi:10.1371/journal.pone.0169748.

105. Waikar, M.L. Nilawar, A.P. *Identification of Groundwater Potential Zone Using Remote Sensing and GIS Technique*; 2007; Vol. 3297;.

106. Lehner, B., Verdin, K., & Jarvis, A. Technical Documentation Version 1.0. *USGS Earth Resour. Obs. Sci. Sioux Falls, SD, USA* **2006**.

107. Karro, E.; Indermitte, E.; Saava, A.; Haamer, K.; Marandi, A. Fluoride Occurrence in Publicly Supplied Drinking Water in Estonia. *Environ. Geol.* **2006**, *50*, 389–396, doi:10.1007/s00254-006-0217-1.

108. India-WRIS Available online: https://indiawris.gov.in/wris/#/ (accessed on 12 August 2022).

109. Gong, G.; Mattevada, S.; O'Bryant, S.E. Comparison of the Accuracy of Kriging and IDW Interpolations in Estimating Groundwater Arsenic Concentrations in Texas. *Environ. Res.* **2014**, *130*, 59–69, doi:10.1016/j.envres.2013.12.005.

110. Addison, M.J.; Rivett, M.O.; Phiri, P.; Mleta, P.; Mblame, E.; Wanangwa, G.; Kalin, R.M. Predicting Groundwater Vulnerability to Geogenic Fluoride Risk: A Screening Method for Malawi and an Opportunity for National Policy Redefinition. *Water* **2020**, *12*, 3123, doi:10.3390/w12113123.

111. Salzman, S.A.; Allinson, G.; Stagnitti, F.; Hill, R.J.; Thwaites, L.; Ierodiaconou, D.; Carr, R.; Sherwood, J.; Versace, V. Adsorption and Desorption Characteristics of Fluoride in the Calcareous and Siliceous Sand Sheet Aquifers of South-West Victoria, Australia. *WIT Trans. Ecol. Environ.* **2008**, *111*, 159–174, doi:10.2495/WP080161.

112. Mukherjee, A.; Verma, S.; Gupta, S.; Henke, K.R.; Bhattacharya, P. Influence of Tectonics,

Sedimentation and Aqueous Flow Cycles on the Origin of Global Groundwater Arsenic: Paradigms from Three Continents. *J. Hydrol.* **2014**, *518*, 284–299, doi:10.1016/j.jhydrol.2013.10.044.

113. Mukherjee, A.; Gupta, S.; Coomar, P.; Fryar, A.E.; Guillot, S.; Verma, S.; Bhattacharya, P.; Bundschuh, J.; Charlet, L. Plate Tectonics Influence on Geogenic Arsenic Cycling: From Primary Sources to Global Groundwater Enrichment. *Sci. Total Environ.* **2019**, *683*, 793–807, doi:10.1016/j.scitotenv.2019.04.255.

114. Alvarez, M. del P.; Carol, E. Geochemical Occurrence of Arsenic, Vanadium and Fluoride in Groundwater of Patagonia, Argentina: Sources and Mobilization Processes. *J. South Am. Earth Sci.* **2019**, *89*, 1–9, doi:10.1016/j.jsames.2018.10.006.

115. Parvez, I.A.; Ram, A. Probabilistic Assessment of Earthquake Hazards in the Indian Subcontinent. *Pure Appl. Geophys.* **1999**, *154*, 23–40, doi:10.1007/s000240050219.

116. Jain, A.K.; Banerjee, D.M.; Kale, V.S. *Tectonics of the Indian Subcontinent*; Society of Earth Scientists Series; Springer International Publishing: Cham, 2020; ISBN 978-3-030-42844-0.

117. Siebert, S.; Henrich, V.; Frenken, K.; Friedrich-Wilhelms-University, J.B.-R.; 2013, undefined Global Map of Irrigation Areas Version 5.