Supplementary Material for "Groundwater chemistry inference, transferability and physical consistency evaluation with machine learning across global monitoring records"

Md Rakib Hasan, Shoumik Zubyer, Anwar Zahid

S1 Data sources and harmonization

The pretraining corpus combines GEMStat v3 (943,058 samples), French ADES (693,720), EEA Waterbase (288,319) and USGS NWIS through the Water Quality Portal (162,873), for 2,087,970 samples in total. Records were mapped to 20 parameters, with dissolved fractions preferred over total ones and physically impossible values removed. The 18 environmental rasters were extracted at 111,791 unique site coordinates through Google Earth Engine (Gorelick et al., 2017). Terrain supplies elevation and slope (NASADEM, 30 m), soil supplies sand and clay fraction, organic carbon, soil pH and bulk density (OpenLandMap, 250 m), land surface supplies land cover (ESA WorldCover, 10 m), population density (WorldPop, 1 km) and the mean and standard deviation of NDVI (MODIS, 500 m), climate supplies precipitation, temperature, actual and potential evapotranspiration and aridity (TerraClimate), and GRACE supplies the mean and trend of the terrestrial water storage anomaly. The Bangladesh dataset of 2,788 samples from 1,507 sites combines 1,807 samples from the authors' 2020–2021 campaign with 981 from the 2012–2013 Bangladesh Water Development Board campaign, and it was excluded from every stage of training.

S2 GENESIS architecture and pretraining

Each measured parameter becomes a token whose embedding sums a parameter-type vector, a two-layer value projection with GELU activation specific to that parameter, and a water-body-type vector. The type vocabulary covers 20 chemistry tokens, 18 raster tokens and 3 special tokens, and the maximum sequence length is 44. Eighteen of the 20 chemistry parameters were transformed as $\log(1 + \max(x, 0))$ before z-scoring, while pH and Eh were z-scored directly, and rasters were z-scored without transformation. The encoder is a Transformer (Vaswani et al., 2017) with pre-layer normalization (Xiong et al., 2020) and a GELU feed-forward network. The Base configuration used in every analysis has 8 layers, 8 attention heads and a width of 256, giving 8.99 million parameters.

Pretraining masks a random 20% of chemistry tokens, replacing them with a mask token and zeroing their values, and minimizes the mean squared error of the reconstructed z-scored values at the masked positions. Rasters are never masked. Training ran in two stages, the first with mask ratio 0.20 for up to 100 epochs and the second with mask ratio 0.25 for up to 50 epochs, warm-started from the best first-stage checkpoint. AdamW (Loshchilov and Hutter, 2019) was used with weight decay 0.01, linear warm-up followed by decay to zero, gradient clipping at 1.0, a first-stage learning rate of 1×10^{-4} and a second-stage rate of 5×10^{-5}, with early stopping on validation loss (patience 20 and 15). The second-stage checkpoint with the lowest validation loss was used throughout.

S3 Inferability

Tables S1 and S2 give R^2 for every target at levels L2 (field parameters and major ions) and L4 (every other parameter), for the five regressors including the simulated quantum kernel (QK), on the site-disjoint wells and in Bangladesh. The best value at each level is in bold. Table S3 gives the greedy measurement sets, selected on validation samples from sites that also appear in training and scored on the site-disjoint wells.

Table S1: R^2 per target at unseen (site-disjoint) groundwater wells, at level L2 (field parameters and major ions plus rasters) and level L4 (every other parameter plus rasters). GB, gradient boosting. The best value at each level is in bold. Eh had no site-disjoint test rows.

Target	n	GENESIS L2	GB L2	RF L2	TabPFN L2	QK L2	GENESIS L4	GB L4	RF L4	TabPFN L4	QK L4
Ca	333	**0.95**	0.94	0.92	0.85	0.63	**0.96**	0.93	0.92	0.88	0.59
Mg	333	**0.87**	0.86	0.82	0.85	0.51	**0.90**	0.86	0.82	0.85	0.50
Na	333	**0.93**	0.92	0.90	0.92	0.68	**0.95**	0.92	0.88	0.92	0.66
K	322	**0.73**	0.64	0.62	0.62	0.43	**0.77**	0.61	0.62	0.61	0.42
Cl	333	**0.90**	0.85	0.80	0.85	0.61	**0.91**	0.85	0.79	0.83	0.60
HCO_3	333	**0.92**	0.91	0.88	0.75	0.70	**0.95**	0.92	0.89	0.71	0.66
SO_4	333	**0.83**	0.78	0.74	0.74	0.57	**0.86**	0.79	0.75	0.75	0.57
EC	196	0.91	0.91	0.90	**0.93**	0.70	0.92	0.91	0.87	**0.94**	0.75
TDS	253	0.93	0.96	0.95	**0.98**	0.93	0.94	0.96	0.95	**0.99**	0.91
pH	146	0.61	0.65	**0.67**	0.66	0.34	**0.77**	0.68	0.68	0.67	0.21
SiO_2	269	**0.73**	0.70	0.70	0.71	0.26	**0.80**	0.73	0.71	0.72	0.25
F	225	**0.78**	0.71	0.70	0.60	0.36	**0.80**	0.71	0.67	0.62	0.37
NO_3	215	**0.65**	0.49	0.44	0.52	0.09	**0.75**	0.58	0.50	0.56	0.21
PO_4	149	**0.81**	0.61	0.57	0.28	0.06	**0.91**	0.89	0.81	0.37	-0.01
DOC	173	0.16	0.28	**0.39**	0.21	0.11	0.39	**0.43**	0.37	0.29	-0.05
Fe	225	**0.53**	0.32	0.26	0.15	0.10	**0.71**	0.63	0.53	0.34	0.22
Mn	249	**0.47**	0.31	0.13	0.10	-0.07	**0.68**	0.49	0.33	0.20	0.30
U	85	0.41	0.53	0.56	**0.61**	0.35	0.64	0.68	0.72	**0.75**	0.57
As	165	0.22	0.37	**0.38**	0.33	0.03	**0.53**	0.51	0.47	0.39	0.12

Table S2: R^2 per target in the Bangladesh held-out set, at levels L2 and L4, as in Table S1. U and DOC were too sparse in Bangladesh to score.

Target	n	GENESIS L2	GB L2	RF L2	TabPFN L2	QK L2	GENESIS L4	GB L4	RF L4	TabPFN L4	QK L4
Ca	2710	**0.42**	0.33	0.31	-0.17	0.14	**0.38**	0.35	0.33	0.11	-0.46
Mg	2710	**0.45**	0.28	0.09	0.25	0.31	**0.46**	0.31	0.18	0.27	0.15
Na	2710	0.47	0.76	**0.77**	0.76	-0.03	0.75	0.77	**0.78**	0.77	0.44
K	2710	0.08	0.11	**0.12**	0.11	-0.43	**0.19**	0.07	0.13	0.17	-0.18
Cl	2710	0.52	0.70	0.70	**0.75**	0.01	0.69	0.71	0.71	**0.78**	0.25
HCO_3	2710	0.21	0.33	**0.34**	-0.10	-0.75	0.15	0.07	**0.36**	-0.06	-0.14
SO_4	2710	**-0.58**	-1.17	-1.73	-1.74	-2.07	-1.10	**-1.07**	-1.70	-1.52	-2.56
EC	2705	0.67	0.79	0.75	**0.81**	-0.58	**0.84**	0.76	0.59	0.82	0.66
TDS	2706	0.76	0.81	0.82	**0.87**	0.76	0.81	0.82	0.82	**0.87**	0.48
pH	2699	**-0.06**	-0.37	-0.19	-0.10	-0.15	-0.20	-0.62	-0.24	**-0.12**	-0.52
SiO_2	419	-2.01	-0.65	-0.23	**-0.20**	-0.38	-0.24	-0.56	**-0.19**	-0.34	-0.41
F	840	-0.25	-0.16	**-0.02**	-0.04	-0.31	-0.11	-0.11	**0.02**	-0.05	-0.29
NO_3	2674	-1.01	-1.10	-0.07	**0.03**	-0.40	-1.65	-1.46	-0.16	**-0.10**	-0.49
PO_4	2672	**-0.53**	-0.71	-0.58	-1.26	-1.34	-0.62	-0.30	**-0.12**	-1.08	-1.34
Fe	2676	-1.64	**-0.79**	-0.89	-2.14	-2.01	-0.76	-0.72	**-0.48**	-1.47	-1.05
Mn	2707	-0.45	-0.14	**-0.10**	-0.40	-0.70	-0.21	-0.30	-0.20	**-0.08**	-0.26
As	2653	-0.32	-0.17	**-0.15**	-0.24	-0.15	-0.25	-0.06	-0.04	0.00	**0.07**
Eh	2665	**-2.17**	-3.15	-3.27	-5.01	-16.11	-1.29	**-0.88**	-4.18	-1.67	-15.82

Table S3: Greedy minimum measurement sets. Starting from pH and EC, the parameter that most raised GENESIS reconstruction R^2 was added at each step. R^2 after each addition is given for GENESIS and for gradient boosting refitted on the same inputs, on the site-disjoint wells.

Target	n	pH+EC	Added in order	GENESIS R^2 per step	GB R^2 per step
As	165	0.09	Fe → K → Mn → Na	0.11 / 0.13 / 0.09 / 0.11	0.31 / 0.38 / 0.33 / 0.34
Fe	225	0.10	Mn → HCO_3 → K → NO_3	0.40 / 0.42 / 0.46 / 0.48	0.55 / 0.60 / 0.54 / 0.54
Mn	249	0.16	Fe → Cl → NO_3 → Na	0.52 / 0.45 / 0.51 / 0.59	0.37 / 0.58 / 0.43 / 0.47
U	85	-0.66	HCO_3 → DOC → Ca → SO_4	0.09 / 0.13 / 0.30 / 0.22	0.60 / 0.68 / 0.65 / 0.66
F	225	-0.02	HCO_3 → Na → NO_3 → Ca	0.35 / 0.57 / 0.61 / 0.61	0.52 / 0.65 / 0.66 / 0.66
NO_3	215	0.14	SO_4 → HCO_3 → Ca → Mn	0.32 / 0.48 / 0.47 / 0.53	0.42 / 0.41 / 0.47 / 0.53
PO_4	149	-0.41	K → Mg → NO_3 → F	0.18 / 0.48 / 0.55 / 0.57	0.64 / 0.63 / 0.71 / 0.85

S4 Transfer and the raster-only audit

Table S4 gives the out-of-region AUC of every method in each of the 47 leave-one-region-out cells, averaged over three seeds. Region labels prefixed G: are GEMStat country sets, and EEA is the EEA Waterbase set.

The audit of raster-only models (Table S5) ranks parameters by the masked-reconstruction R^2 of GENESIS on the pretraining corpus. That reconstruction score was computed on samples drawn from the corpus at the sample level, so it is not site-disjoint and overstates absolute accuracy at unseen sites. It is used here only to rank parameters by how well chemistry determines them, and that ranking agrees with the site-disjoint L4 values of Table S1 (Spearman $\rho = 0.81$ over the 19 parameters scored in both). Labels mark values above each parameter's 80th percentile of site medians, so every parameter has 20% positives, and each repeat draws 3,000 sites. Table S6 gives the controls. Within single held-out regions the ordering held in the USA, France, the EEA set and Bangladesh (ρ 0.66–0.78), was weak in Greece and absent in Mexico, while within one data source it held for the French network and not for the US network. The guideline-threshold variant of the audit, with prevalence free to vary, gave the same pattern of high random-split AUC and lower leave-one-region-out AUC for all seven contaminants (As 0.75 to 0.55, Mn 0.79 to 0.55).

Table S4: Out-of-region exceedance AUC per leave-one-region-out cell, mean of three seeds. n is the number of held-out pairs and Exc. the number of exceedances among them. GEN-FT, GENESIS fine-tuned. GEN-Fr, GENESIS frozen probe. Raster, raster-only random forest. kNN, spatial interpolation from the 25 nearest training sites.

Target	Region	n	Exc.	GEN-FT	GEN-Fr	TabPFN	XGB	RF	LogReg	Raster	kNN
As	Bangladesh	224	70	0.64	0.61	0.61	0.62	0.68	0.56	0.50	0.50
As	EEA	3969	106	0.75	0.67	0.71	0.63	0.65	0.64	0.62	0.47
As	France	23226	695	0.79	0.77	0.79	0.76	0.75	0.80	0.53	0.49
As	G:Italy	1314	46	0.76	0.72	0.76	0.71	0.69	0.67	0.55	0.74
As	G:Mexico	3509	532	0.74	0.75	0.74	0.72	0.74	0.70	0.46	0.60
As	G:Poland	123	14	0.41	0.41	0.27	0.35	0.26	0.50	0.53	0.54
As	USA	6290	782	0.90	0.84	0.91	0.86	0.91	0.81	0.56	0.47
Fe	Bangladesh	224	212	0.61	0.58	0.67	0.55	0.47	0.72	0.51	0.52
Fe	EEA	2451	554	0.82	0.82	0.84	0.82	0.83	0.83	0.72	0.74
Fe	France	25570	1708	0.85	0.82	0.85	0.84	0.83	0.76	0.59	0.62
Fe	G:France	468	65	0.74	0.76	0.78	0.76	0.78	0.62	0.68	0.64
Fe	G:India	148	52	0.54	0.44	0.54	0.61	0.49	0.44	0.54	0.49
Fe	G:Ireland	121	25	0.69	0.76	0.73	0.64	0.67	0.73	0.60	0.24
Fe	G:Italy	457	14	0.71	0.77	0.71	0.73	0.66	0.71	0.57	0.36
Fe	G:Mexico	1159	83	0.68	0.65	0.70	0.71	0.72	0.50	0.46	0.58
Fe	G:Sweden	219	110	0.93	0.88	0.93	0.93	0.84	0.90	0.89	0.73
Fe	USA	5038	1657	0.86	0.81	0.87	0.84	0.87	0.83	0.60	0.59
Mn	Bangladesh	224	96	0.55	0.54	0.57	0.57	0.58	0.55	0.46	0.50
Mn	EEA	2516	30	0.56	0.52	0.66	0.59	0.61	0.64	0.57	0.58
Mn	France	25456	574	0.87	0.85	0.83	0.83	0.80	0.81	0.61	0.51

Table S4: Out-of-region exceedance AUC per leave-one-region-out cell, mean of three seeds. n is the number of held-out pairs and Exc. the number of exceedances among them. GEN-FT, GENESIS fine-tuned. GEN-Fr, GENESIS frozen probe. Raster, raster-only random forest. kNN, spatial interpolation from the 25 nearest training sites.

Target	Region	n	Exc.	GEN-FT	GEN-Fr	TabPFN	XGB	RF	LogReg	Raster	kNN
Mn	G:Mexico	1257	38	0.86	0.81	0.88	0.86	0.84	0.81	0.56	0.59
Mn	G:Netherlands	106	16	0.42	0.41	0.46	0.32	0.45	0.62	0.39	0.50
Mn	USA	6381	756	0.91	0.82	0.93	0.91	0.93	0.90	0.53	0.47
PO_4	Bangladesh	224	185	0.68	0.71	0.65	0.61	0.58	0.61	0.54	0.49
PO_4	EEA	6385	1389	0.72	0.65	0.69	0.65	0.70	0.72	0.71	0.48
PO_4	France	7011	187	0.73	0.69	0.73	0.67	0.69	0.64	0.57	0.62
PO_4	G:Greece	112	16	0.91	0.86	0.89	0.89	0.87	0.89	0.78	0.67
PO_4	G:India	327	38	0.72	0.69	0.61	0.64	0.65	0.43	0.82	0.66
PO_4	G:Italy	1377	69	0.73	0.73	0.67	0.68	0.65	0.57	0.72	0.61
PO_4	G:Lithuania	339	30	0.88	0.84	0.83	0.79	0.73	0.65	0.72	0.51
PO_4	G:Mexico	4009	948	0.84	0.78	0.83	0.81	0.78	0.83	0.55	0.45
PO_4	G:Netherlands	313	20	0.93	0.93	0.98	0.81	0.94	0.96	0.44	0.84
PO_4	G:Poland	1182	47	0.75	0.73	0.71	0.71	0.72	0.58	0.64	0.49
PO_4	USA	5596	93	0.82	0.78	0.86	0.78	0.88	0.79	0.53	0.59
NO_3	EEA	6342	145	0.95	0.91	0.95	0.93	0.94	0.93	0.74	0.44
NO_3	France	27372	1433	0.90	0.81	0.93	0.90	0.90	0.91	0.62	0.55
NO_3	G:Greece	598	75	0.80	0.76	0.77	0.73	0.70	0.77	0.72	0.50
NO_3	G:Mexico	4035	12	0.82	0.89	0.88	0.88	0.91	0.89	0.63	0.49
NO_3	USA	5180	18	0.85	0.78	0.94	0.87	0.92	0.98	0.57	0.50
F	EEA	1630	18	0.75	0.59	0.80	0.72	0.66	0.46	0.65	0.59
F	France	2563	60	0.88	0.86	0.89	0.87	0.87	0.87	0.48	0.40
F	G:Mexico	998	204	0.85	0.81	0.82	0.84	0.84	0.80	0.62	0.44
F	USA	8085	544	0.96	0.89	0.96	0.97	0.96	0.93	0.73	0.66
U	France	1691	14	0.57	0.58	0.92	0.93	0.91	0.95	0.09	0.09
U	G:Canada	209	138	0.51	0.53	0.51	0.55	0.55	0.51	0.50	0.34
U	G:France	351	34	0.65	0.50	0.33	0.46	0.33	0.37	0.38	0.40
U	USA	3792	171	0.70	0.67	0.81	0.55	0.62	0.33	0.65	0.70
Mean of 47 cells				0.75	0.72	0.76	0.73	0.73	0.71	0.58	0.53

Table S5: Raster-only random forest AUC per parameter under random cross-validation, spatial block cross-validation at 1°, 5° and 10°, and leave-one-region-out validation (LORO), mean of five repeats of 3,000 sites. Gap is random minus LORO AUC. Parameters are ordered by masked-reconstruction R^2, which is computed on the corpus and is not site-disjoint.

Parameter	Recon. R^2	Random	Block 1°	Block 5°	Block 10°	LORO	Gap
Eh	0.59	0.83	0.74	0.67	0.61	0.37	0.46
As	0.67	0.74	0.70	0.67	0.66	0.58	0.16
U	0.70	0.85	0.81	0.78	0.75	0.57	0.29
Mn	0.71	0.78	0.76	0.72	0.68	0.54	0.24
PO_4	0.72	0.79	0.70	0.63	0.60	0.54	0.25
Fe	0.75	0.79	0.75	0.74	0.71	0.56	0.23
pH	0.78	0.77	0.75	0.72	0.69	0.58	0.19
F	0.78	0.82	0.80	0.78	0.77	0.57	0.25
DOC	0.78	0.79	0.74	0.70	0.67	0.48	0.31
NO_3	0.80	0.86	0.83	0.81	0.78	0.62	0.23
K	0.84	0.79	0.77	0.73	0.72	0.58	0.21
SiO_2	0.84	0.89	0.86	0.80	0.73	0.52	0.37
SO_4	0.88	0.83	0.80	0.77	0.75	0.68	0.15
Mg	0.91	0.81	0.78	0.75	0.72	0.62	0.20
Cl	0.93	0.82	0.79	0.75	0.73	0.71	0.11
HCO_3	0.93	0.77	0.74	0.71	0.67	0.63	0.15
Na	0.94	0.84	0.82	0.79	0.77	0.69	0.15
Ca	0.95	0.80	0.78	0.76	0.74	0.66	0.14
TDS	0.95	0.82	0.77	0.73	0.72	0.68	0.14
EC	0.96	0.83	0.79	0.76	0.75	0.67	0.15

Table S6: Controls for the raster-only transfer gradient. Spearman correlation between masked-reconstruction R^2 and raster-only AUC across parameters, within each held-out region with at least six scorable parameters, and within a single data source under spatial block hold-out. Across-parameter and within-source p values are from 10,000 permutations.

Control	Scope	Spearman ρ	p	Parameters
Across parameters, random CV	all regions	0.20	0.4044	20
Across parameters, LORO	all regions	0.76	0.0001	20
Within held-out region	Bangladesh	0.69	0.002	17
Within held-out region	EEA	0.66	0.004	17
Within held-out region	France	0.66	0.002	20
Within held-out region	G:Greece	0.14	0.787	6
Within held-out region	G:Mexico	0.03	0.911	17
Within held-out region	USA	0.78	0.000	20
Within one source, block hold-out	France	0.77	0.0001	20
Within one source, block hold-out	USA	0.31	0.1888	20

S5 Calibration

Table S7 gives the out-of-region coverage of the exceedance class per cell for every method, calibrated by class-conditional split conformal prediction at $\alpha = 0.10$ on out-of-region pairs. Table S8 gives the effect of recalibrating on k labelled pairs from the held-out region. The values in parentheses are the fraction of samples whose prediction set contains both classes, which shows that the restored coverage comes from ambiguous sets.

Table S7: Out-of-region coverage of the exceedance class per leave-one-region-out cell (nominal 0.90), mean of three seeds. Abbreviations as in Table S4.

Target	Region	n	Exc.	GEN-FT	GEN-Fr	TabPFN	XGB	RF	LogReg	Raster	kNN
As	Bangladesh	224	70	0.75	0.66	1.00	0.61	1.00	0.96	0.95	1.00
As	EEA	3969	106	0.82	0.61	0.81	0.72	0.69	0.78	0.97	0.46
As	France	23226	695	0.52	0.48	0.69	0.65	0.68	0.79	0.06	0.02
As	G:Italy	1314	46	0.66	0.58	0.71	0.80	0.72	0.76	0.80	0.68
As	G:Mexico	3509	532	0.68	0.82	0.82	0.64	0.93	0.99	1.00	0.58
As	G:Poland	123	14	0.31	0.71	0.43	0.67	0.24	0.95	0.86	0.48
As	USA	6290	782	1.00	0.99	0.99	0.91	0.99	0.99	1.00	0.89
Fe	Bangladesh	224	212	0.99	0.98	1.00	0.89	1.00	0.97	1.00	1.00
Fe	EEA	2451	554	0.84	0.80	0.85	0.80	0.85	0.91	0.94	0.66
Fe	France	25570	1708	0.64	0.65	0.72	0.73	0.78	0.84	0.48	0.45
Fe	G:France	468	65	0.82	0.84	0.96	0.95	0.86	0.54	0.76	0.71
Fe	G:India	148	52	0.48	0.60	0.93	0.75	0.68	0.67	0.85	1.00
Fe	G:Ireland	121	25	1.00	1.00	1.00	0.99	0.97	0.88	1.00	1.00
Fe	G:Italy	457	14	0.55	0.76	0.74	0.86	0.64	0.81	0.81	0.00
Fe	G:Mexico	1159	83	0.51	0.58	0.77	0.58	0.80	0.79	0.98	0.98
Fe	G:Sweden	219	110	0.99	0.99	0.99	0.99	0.89	0.98	1.00	1.00
Fe	USA	5038	1657	0.94	0.97	0.97	0.89	0.97	0.97	0.95	0.85
Mn	Bangladesh	224	96	0.88	0.79	0.98	0.69	0.99	1.00	0.89	0.67
Mn	EEA	2516	30	0.83	0.52	0.89	0.84	0.78	0.63	0.92	1.00
Mn	France	25456	574	0.70	0.67	0.72	0.72	0.71	0.80	0.40	0.02
Mn	G:Mexico	1257	38	0.62	0.62	0.96	0.74	0.89	0.89	1.00	0.56
Mn	G:Netherlands	106	16	0.60	0.81	0.85	0.77	0.83	0.85	0.90	1.00
Mn	USA	6381	756	0.96	0.94	0.97	0.92	0.98	0.95	0.98	1.00
PO_4	Bangladesh	224	185	0.96	0.96	0.99	0.69	0.98	0.90	0.99	1.00
PO_4	EEA	6385	1389	0.68	0.66	0.72	0.74	0.76	0.92	0.76	0.38
PO_4	France	7011	187	0.27	0.34	0.70	0.46	0.70	0.75	0.34	0.34
PO_4	G:Greece	112	16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
PO_4	G:India	327	38	0.39	0.64	0.24	0.11	0.33	0.47	0.99	1.00
PO_4	G:Italy	1377	69	0.51	0.78	0.36	0.32	0.44	0.75	0.90	0.76
PO_4	G:Lithuania	339	30	0.07	0.00	0.67	0.60	0.69	0.97	0.70	0.68
PO_4	G:Mexico	4009	948	0.76	0.81	0.84	0.69	0.90	0.88	0.98	0.68
PO_4	G:Netherlands	313	20	0.87	0.93	1.00	0.85	1.00	1.00	0.38	0.95
PO_4	G:Poland	1182	47	0.73	0.81	0.91	0.79	0.85	0.89	0.81	0.56

Table S7: Out-of-region coverage of the exceedance class per leave-one-region-out cell (nominal 0.90), mean of three seeds. Abbreviations as in Table S4.

Target	Region	n	Exc.	GEN-FT	GEN-Fr	TabPFN	XGB	RF	LogReg	Raster	kNN
PO_4	USA	5596	93	0.53	0.70	0.92	0.55	0.95	0.86	0.97	0.59
NO_3	EEA	6342	145	0.91	0.75	0.92	0.90	0.93	0.92	0.88	0.09
NO_3	France	27372	1433	0.90	0.92	0.98	0.94	0.97	0.97	1.00	1.00
NO_3	G:Greece	598	75	0.27	0.32	0.27	0.26	0.33	0.24	0.92	1.00
NO_3	G:Mexico	4035	12	0.53	0.53	0.67	0.22	0.78	0.64	0.61	0.67
NO_3	USA	5180	18	0.63	0.61	0.81	0.44	0.80	0.83	0.83	0.67
F	EEA	1630	18	0.17	0.15	0.37	0.31	0.30	0.46	0.93	0.91
F	France	2563	60	0.55	0.60	0.62	0.70	0.63	0.74	0.08	0.41
F	G:Mexico	998	204	0.76	0.87	0.79	0.74	0.87	0.86	1.00	1.00
F	USA	8085	544	0.98	0.93	0.96	0.96	0.97	0.97	0.90	0.89
U	France	1691	14	0.29	0.45	0.93	0.93	0.93	1.00	0.31	0.00
U	G:Canada	209	138	0.47	0.45	0.42	0.27	0.45	0.16	1.00	1.00
U	G:France	351	34	0.00	0.20	0.07	0.30	0.04	0.29	0.05	0.67
U	USA	3792	171	0.34	0.73	0.46	0.38	0.51	0.23	1.00	1.00
Mean of 47 cells				0.65	0.69	0.77	0.69	0.77	0.80	0.80	0.71

Table S8: Exceedance coverage in the held-out region before and after recalibration on k labelled pairs from that region (50 draws each), averaged over cells. Values in parentheses are the fraction of ambiguous prediction sets.

Method	No recal.	$k = 25$	$k = 50$	$k = 100$	$k = 200$
GEN-FT	0.65	0.99 (0.87)	0.98 (0.81)	0.96 (0.74)	0.95 (0.67)
GEN-Fr	0.69	0.99 (0.88)	0.98 (0.83)	0.96 (0.77)	0.95 (0.70)
TabPFN	0.77	0.99 (0.87)	0.98 (0.82)	0.96 (0.76)	0.95 (0.68)
XGB	0.69	0.99 (0.88)	0.98 (0.82)	0.96 (0.76)	0.95 (0.71)
RF	0.77	0.99 (0.88)	0.98 (0.83)	0.96 (0.77)	0.95 (0.71)
LogReg	0.80	0.99 (0.88)	0.98 (0.83)	0.96 (0.79)	0.95 (0.74)
Raster	0.80	0.99 (0.90)	0.98 (0.87)	0.96 (0.83)	0.95 (0.81)
kNN	0.71	0.99 (0.95)	0.98 (0.93)	0.96 (0.90)	0.95 (0.88)

S6 Physical consistency

Table S9 gives saturation-index agreement and charge-balance error per regressor, test set and redox class. Saturation indices used Davies activity coefficients at 25 °C and free-ion concentrations without ion pairs, which suffices for comparing reconstructed with observed waters computed by identical equations but not for absolute saturation states in saline or strongly complexing waters. Cal. state is the fraction of samples whose calcite saturation state (over- or undersaturated) is the same in reconstructed and observed water.

S7 Simulated quantum kernel

Table S10 compares the quantum kernel with every other classifier on out-of-region AUC and exceedance coverage, cell by cell. The circuit applies, twice, a layer of Hadamard gates followed by the diagonal unitary $\exp(i \sum_k a_k Z_k + i \sum_k a_k a_{k+1} Z_k Z_{k+1})$ over a ring of 12 qubits, where $a = cz$, z holds the 12 standardized principal components clipped to ± 3 and c is the bandwidth. The control without entanglement drops the ZZ terms. The kernel is the squared overlap of the two statevectors, computed exactly in single precision, and the simulator was checked against dense matrix products on four qubits. A fit took a median 32 s per cell on an Apple M1 CPU for either circuit and 2 s for the RBF kernel. Over the 141 cell and seed fits the selected bandwidth was 0.01 in 5, 0.02 in 7, 0.05 in 26, 0.1 in 48, 0.2 in 48 and 0.4 in 7, and the median variance of off-diagonal training kernel values was 0.015 for the quantum kernel, 0.015 for the circuit without entanglement and 0.041 for the RBF kernel.

Table S9: Physical consistency of L4 reconstructions. R^2 of saturation indices computed from reconstructed against observed waters, agreement in calcite saturation state, and median absolute charge-balance error (%) of reconstructed and observed waters. Reducing samples have Fe $\geq$ 0.1 mg/L or Mn $\geq$ 0.05 mg/L.

Test set	Method	n	Calcite	Gypsum	Fluorite	Cal. state	|CBE| rec.	|CBE| obs.
Unseen sites, all	GENESIS	142	0.91	0.94	0.41	0.85	4.4	3.4
	GB	142	0.90	0.92	0.34	0.85	5.8	3.4
	RF	142	0.88	0.89	0.36	0.84	4.2	3.4
	TabPFN	142	0.87	0.89	0.41	0.83	6.1	3.4
	QK	142	0.74	0.75	-1.84	0.73	8.1	3.4
Unseen sites, oxic	GENESIS	94	0.90	0.95	0.34	0.85	4.8	3.2
	GB	94	0.88	0.95	0.29	0.85	5.6	3.2
	RF	94	0.86	0.91	0.30	0.81	3.9	3.2
	TabPFN	94	0.87	0.92	0.36	0.81	4.8	3.2
	QK	94	0.74	0.79	-2.40	0.71	8.3	3.2
Unseen sites, reducing	GENESIS	48	0.93	0.90	0.31	0.83	4.1	3.6
	GB	48	0.93	0.83	0.01	0.85	6.9	3.6
	RF	48	0.90	0.81	0.17	0.90	5.1	3.6
	TabPFN	48	0.86	0.81	0.14	0.88	7.8	3.6
	QK	48	0.73	0.64	-0.29	0.77	7.8	3.6
Bangladesh	GENESIS	2000	0.16	-0.43	-0.07	0.64	10.9	7.5
	GB	2000	-0.34	-0.40	-0.21	0.60	19.7	7.5
	RF	2000	0.15	-0.57	-0.20	0.56	15.3	7.5
	TabPFN	2000	-0.07	-0.54	-0.13	0.61	13.4	7.5
	QK	2000	0.14	-0.68	-1.08	0.65	31.3	7.5

Table S10: Simulated quantum kernel (QK) against each classifier, paired by leave-one-region-out cell after averaging three seeds. AUC and exceedance coverage (Cov.) in the held-out region, the number of cells in which the quantum kernel had the higher AUC, and two-sided Wilcoxon p values. RBF kernel and No entanglement are the matched controls on identical rows and features. Uranium (four cells) is included in All only.

Compared with	Cells	n	AUC QK	AUC other	QK higher	p	Cov. QK	Cov. other	p
RBF kernel	All	47	0.708	0.708	26	0.6558	0.802	0.803	0.8171
	Redox	34	0.694	0.696	19	0.8132	0.842	0.842	0.8204
	Conservative	9	0.831	0.824	5	0.5703	0.737	0.722	0.4961
No entanglement	All	47	0.708	0.708	25	0.9041	0.802	0.801	0.9436
	Redox	34	0.694	0.695	18	0.9328	0.842	0.841	0.8612
	Conservative	9	0.831	0.823	5	0.4258	0.737	0.718	0.1562
GEN-FT	All	47	0.708	0.755	13	<0.0001	0.802	0.652	<0.0001
	Redox	34	0.694	0.743	9	0.0003	0.842	0.702	0.0002
	Conservative	9	0.831	0.862	3	0.0977	0.737	0.632	0.0977
TabPFN	All	47	0.708	0.760	6	<0.0001	0.802	0.774	0.2718
	Redox	34	0.694	0.742	6	0.0001	0.842	0.827	0.7243
	Conservative	9	0.831	0.882	0	0.0039	0.737	0.710	0.6523
XGB	All	47	0.708	0.732	18	0.0398	0.802	0.687	<0.0001
	Redox	34	0.694	0.712	14	0.2214	0.842	0.733	0.0005
	Conservative	9	0.831	0.856	3	0.1641	0.737	0.609	0.0273
RF	All	47	0.708	0.731	17	0.0264	0.802	0.765	0.1368
	Redox	34	0.694	0.713	15	0.2087	0.842	0.808	0.2865
	Conservative	9	0.831	0.857	2	0.1641	0.737	0.730	1.0000
GEN-Fr	All	47	0.708	0.723	20	0.2005	0.802	0.692	<0.0001
	Redox	34	0.694	0.717	12	0.0696	0.842	0.736	0.0011
	Conservative	9	0.831	0.812	7	0.1289	0.737	0.630	0.0547
LogReg	All	47	0.708	0.711	22	0.5963	0.802	0.796	0.8979
	Redox	34	0.694	0.698	17	0.9731	0.842	0.855	0.9375
	Conservative	9	0.831	0.836	4	0.5703	0.737	0.739	0.8203
Raster	All	47	0.708	0.583	35	<0.0001	0.802	0.805	0.2861
	Redox	34	0.694	0.590	26	0.0002	0.842	0.833	0.4265
	Conservative	9	0.831	0.640	8	0.0078	0.737	0.794	0.3828
kNN	All	47	0.708	0.532	39	<0.0001	0.802	0.707	0.1095
	Redox	34	0.694	0.555	28	<0.0001	0.842	0.704	0.0098
	Conservative	9	0.831	0.509	9	0.0039	0.737	0.737	0.6523

For inferability and physical consistency the same three kernels were used in kernel ridge regression (main-text Methods). Table S11 gives the median R^2 over targets at each input level for the three kernels and for gradient boosting, with paired Wilcoxon tests of the quantum kernel against each control, and Table S12 gives the physical consistency of their L4 reconstructions. The chosen ridge penalty lay at the upper edge of its range in 2% of fits.

Table S11: Median R^2 over targets for kernel ridge regression with the quantum kernel (QK), the same circuit without entanglement (No ent.) and the classical RBF kernel, with gradient boosting (GB) for reference, at unseen sites and in Bangladesh. p values are two-sided Wilcoxon tests paired by target.

Test set	Level	Targets	QK	No ent.	RBF	GB	p QK–RBF	p QK–No ent.
Unseen sites	L0	19	0.16	0.16	0.21	0.40	0.055	0.040
	L1	19	0.26	0.21	0.29	0.51	0.984	0.395
	L2	19	0.36	0.40	0.43	0.70	0.293	0.312
	L3	19	0.32	0.46	0.40	0.73	0.490	0.002
	L4	19	0.42	0.46	0.44	0.73	0.293	0.080
Bangladesh	L0	18	-0.77	-0.67	-0.99	-0.94	0.284	0.671
	L1	18	-0.48	-0.59	-0.60	-0.42	0.580	0.468
	L2	18	-0.39	-0.25	-0.33	-0.15	0.865	0.181
	L3	18	-0.35	-0.33	-0.23	-0.17	0.393	0.325
	L4	18	-0.27	-0.22	-0.21	-0.08	0.580	0.417

Table S12: Physical consistency of L4 reconstructions by the three kernels, as in Table S9. |CBE| is the median absolute charge-balance error (%).

Test set	Kernel	n	Calcite	Gypsum	Fluorite	Cal. state	|CBE| rec.	|CBE| obs.
Unseen sites	QK	142	0.74	0.75	-1.84	0.73	8.1	3.4
	No ent.	142	0.74	0.75	-1.88	0.72	8.2	3.4
	RBF	142	0.74	0.76	-0.64	0.75	8.7	3.4
Bangladesh	QK	2000	0.14	-0.68	-1.08	0.65	31.3	7.5
	No ent.	2000	-0.02	-0.71	-1.23	0.61	33.3	7.5
	RBF	2000	0.07	-0.60	-0.84	0.63	15.2	7.5

References

N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, 2017.

I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *Proceedings of ICLR*, 2019.

A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017.

R. Xiong, Y. Yang, J. He, et al. On layer normalization in the Transformer architecture. In *Proceedings of ICML*, pages 10524–10533, 2020.