Supplement of Hydrol. Earth Syst. Sci., 30, 5647–5683, 2026
https://doi.org/10.5194/hess-30-5647-2026-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

Hydrochemistry and modeling nitrate concentration in farmland groundwater under different hydrological seasons by integrating hybrid quantum-classical ML, virtual sample generation and AlphaEarth Foundation

Junjie Xu et al.

Correspondence to: Yilei Yu (yileiyu@hbu.edu.cn)

S1. Groundwater sampling procedures and data quality control

Prior to sample collection, each well was purged by pumping. Sampling commenced only after the pumped volume exceeded three times the well's casing volume and on-site parameters had stabilized (i.e., showing minor fluctuations around a constant value rather than a continuous rising or falling trend), a procedure implemented to ensure the representativeness of the samples. At each sampling point, one 1000 mL and two 100 mL samples were collected. Before final collection, the sample bottles were rinsed three times with the water to be sampled. Immediately after collection, the samples were sealed and stored in a portable cooler for transport to the laboratory for subsequent analysis. Furthermore, the precise geographical location of each sampling point was recorded using a GPS device. Regarding data completeness and missing data handling, all collected samples underwent rigorous screening. Samples with incomplete physicochemical records or those failing initial quality checks were excluded from the final dataset. Consequently, the modeling dataset achieved 100% completeness with no missing values requiring imputation or interpolation.

S2. Analytical methods

In-situ physicochemical parameters, including water temperature (T), pH, total dissolved solids (TDS), dissolved oxygen (DO), electrical conductivity (EC), and oxidation-reduction potential (ORP), were measured using a Hach HQ400 multi-parameter water quality meter. These field-measurable parameters were used as predictors for preliminary nitrate estimation due to their rapid acquisition and minimal sample preparation. HCO_3^- was determined within 24 hours of collection by dilute sulfuric acid-methyl orange titration (Huang et al., 2012). After filtration through 0.45 μm membranes, major cations (K^+, Ca^{2+}, Na^+, Mg^{2+}) were analyzed by ICP-OES (Avio 500) and major anions (NO_3^-, Cl^-, SO_4^{2-}) by ion chromatography (ICS-2100), with analytical precision within ±0.2 mg L^{-1} and charge balance errors below 5%. Nitrite nitrogen and ammonia nitrogen were measured by flow injection analysis (Smartchem 200, AMS Alliance) using dual-wavelength spectrophotometry and the indophenol blue method, respectively (Kim et al., 2019), with detection limits of 0.01 mg L^{-1}. Stable water isotopes (δ^2H, $\delta^{17}O$, $\delta^{18}O$) were analyzed using an LGR liquid water isotope analyzer (TIWA-45-EP) with precisions of ±0.15‰, ±0.02‰, and ±0.02‰, respectively (Hamidi et al., 2023). Nitrate isotopes ($\delta^{18}O$-NO_3^- and $\delta^{15}N$-NO_3^-) were determined by a MAT-253 mass spectrometer coupled with an elemental analyzer, with quality control using international standards (USGS 32, 34, 35), reagent blanks, and duplicate samples. All isotope results are reported in per mil (‰). No interpolation or smoothing was applied to the raw hydrochemical data, except that ordinary Kriging was used in Surfer (Golden Software, LLC) solely for the spatial visualization of nitrate distributions, not for data augmentation, imputation, or model inputs.

References

Kim, J.H., Kang, Y.J., Kim, K.I., Kim, S.K., and Kim, J.H.: Toxic effects of nitrogenous compounds (ammonia, nitrite, and nitrate) on acute toxicity and antioxidant responses of juvenile olive flounder, Paralichthys olivaceus, Environ. Toxicol. Pharmacol., 67, 73–78, https://doi.org/10.1016/j.etap.2019.02.001, 2019.

Hamidi, M.D., Gröcke, D.R., Joshi, S.K., and Greenwell, H.C.: Investigating groundwater recharge using hydrogen and oxygen stable isotopes in Kabul city, a semi-arid region, J. Hydrol., 626, 130187, https://doi.org/10.1016/j.jhydrol.2023.130187, 2023.

S3. AEF data acquisition and preprocessing workflow

The 64-dimensional AEF vectors were spatially sampled at a 10 m resolution using the GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL product on the Google Earth Engine (GEE) platform, based on the geographic coordinates of the field sampling points. To ensure data quality, only samples exhibiting exact matches between the GEE extraction and the actual field data points were retained. Pixels with significant cloud cover or data gaps in the underlying satellite imagery were automatically handled by the AEF pre-processing pipeline, ensuring complete feature vectors for all sampling locations. Given the redundancy within the initial 64-dimensional features, Principal Component Analysis (PCA) based on Singular Value Decomposition (SVD) was employed for feature compression. SVD was performed on the centered feature matrix to select the minimum number of principal components accounting for at least 95% of the cumulative explained variance, balancing information retention with dimensionality reduction to avoid overfitting (Ilyas et al., 2025). The orthogonalized, low-dimensional principal component scores were subsequently used as model inputs, preserving the majority of the semantic information while mitigating overfitting risk.

References

Ilyas, M., Niaz, R., Persio, L.D., Al-Rezami, A.Y., Almazah, M.M., and Tariq, A.: A spatiotemporal approach for drought monitoring using Empirical Orthogonal Function (EOF) analysis and neural networks, Earth Syst. Environ., 10, 2565–2584, https://doi.org/10.1007/s41748-025-00767-z, 2026.

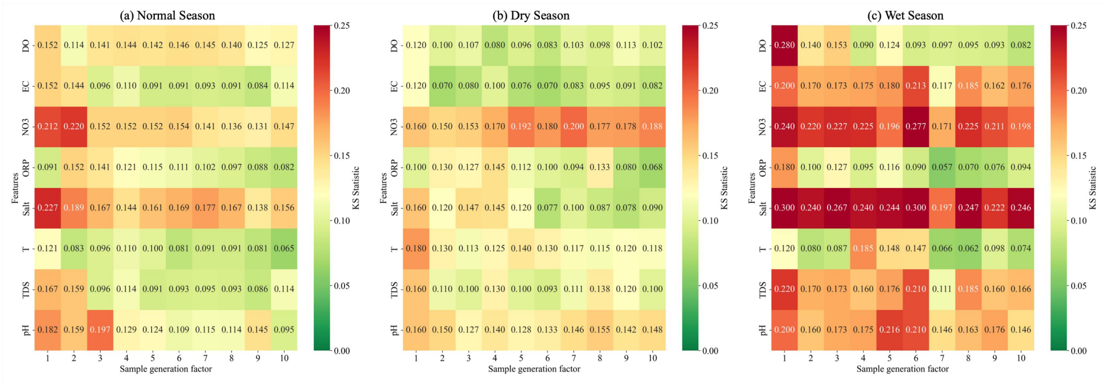

Fig.S1. Heatmaps of KS Statistics for features across different seasons and sample generation factors.

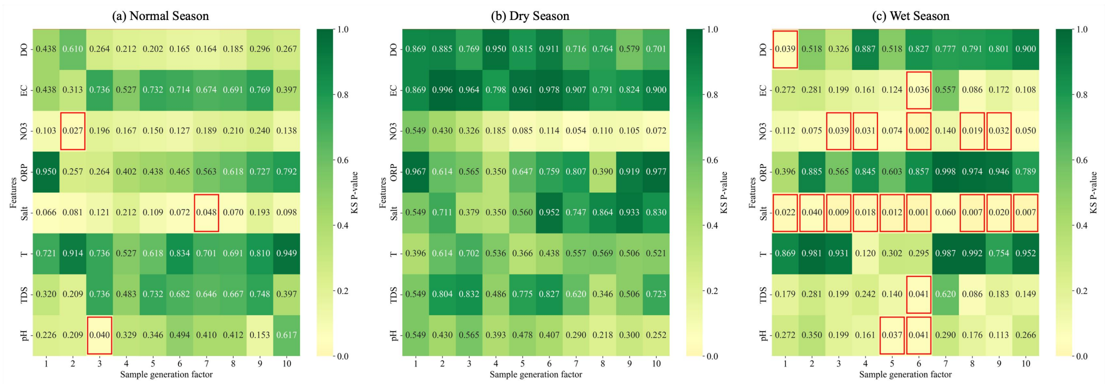

Fig.S2. Heatmaps of KS test P-value for features across different seasons and sample generation factors.

Table S1. Statistical summary of chemical and field measurement parameters.

Periods		pH	T	EC	DO	ORP	Salt	TDS	Depth	K^+	Ca^{2+}
	Unit		℃	μs cm^{-1}	mg L^{-1}	mv	ppt	mg L^{-1}	m	mg L^{-1}	mg L^{-1}
Normal season	Max	8.60	16.70	1110.00	10.57	369.60	0.53	724.00	20.64	36.66	69.73
n=66	Min	7.61	13.40	349.00	2.20	-58.30	0.11	225.00	18.26	1.04	13.38
	Mean	8.16	14.90	549.32	6.31	4.57	0.21	357.21	18.91	2.92	37.28
	SD	0.11	0.71	173.56	1.94	56.27	0.09	113.34	0.62	4.45	13.06
	CV	0.01	0.05	0.32	0.31	12.32	0.44	0.32	0.03	1.52	0.35
Dry season	Max	8.21	18.80	1134.00	8.82	144.20	0.54	737.00	18.95	3.52	43.86
n=65	Min	6.97	14.10	343.00	1.85	-110.40	0.11	227.00	17.88	0.67	4.81
	Mean	7.35	15.59	658.68	6.34	4.55	0.27	427.65	18.43	1.93	16.89
	SD	0.45	0.92	185.20	1.64	42.10	0.10	120.55	0.27	0.56	8.35
	CV	0.06	0.06	0.28	0.26	9.26	0.36	0.28	0.01	0.29	0.49
Wet season	Max	8.97	21.00	977.00	9.61	195.00	0.41	635.00	18.87	3.37	51.54
n=50	Min	5.77	15.50	371.00	3.34	-112.20	0.12	243.00	17.61	1.14	17.27
	Mean	7.34	17.01	535.24	6.98	17.57	0.20	347.92	18.27	1.75	25.70
	SD	0.52	0.94	153.65	1.32	47.59	0.08	100.39	0.40	0.34	6.83
	CV	0.07	0.06	0.29	0.19	2.71	0.42	0.29	0.02	0.19	0.27
		Na^+	Mg^{2+}	HCO_3^-	Cl^-	SO_4^{2-}	F	NO_3^-	NO_2^-	NH_4^+	
	Unit	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	
Normal season	Max	98.85	92.78	192.15	87.09	40.49	13.17	161.17	0.45	0.20	
n=66	Min	3.81	3.80	22.88	1.03	1.79	0.12	2.39	0.04	0.00	
	Mean	24.01	31.78	96.65	20.53	15.53	3.70	33.67	0.11	0.04	
	SD	12.46	17.18	44.51	17.84	10.02	2.02	35.83	0.08	0.04	
	CV	0.52	0.54	0.46	0.87	0.64	0.55	1.06	0.72	1.00	
Dry season	Max	123.20	138.90	289.29	241.36	123.75	0.57	358.58	10.38	0.84	
n=65	Min	16.28	17.97	5.10	1.40	2.00	0.21	0.10	0.57	0.05	
	Mean	48.52	58.17	42.19	24.90	15.91	0.33	42.93	3.35	0.16	
	SD	16.48	25.86	71.11	35.13	18.67	0.08	56.35	1.98	0.11	
	CV	0.34	0.44	1.69	1.41	1.17	0.26	1.31	0.59	0.69	
Wet season	Max	36.21	53.72	207.40	51.93	34.15	0.45	98.36	4.31	0.41	
n=50	Min	10.95	12.76	83.88	1.56	5.88	0.09	4.15	1.51	0.02	
	Mean	21.78	22.57	132.49	15.60	15.16	0.20	27.14	2.62	0.10	

SD	4.90	8.13	27.12	13.60	7.47	0.08	23.86	0.61	0.08
CV	0.23	0.36	0.20	0.87	0.49	0.39	0.88	0.23	0.83

Table S2. Performance evaluation of classical random forest and quantum feature enhanced random forest during the normal season under different numbers of virtual samples.

Sample size	RF				Quantum-enhanced RF			
	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE mg L^{-1}	MAE mg L^{-1}	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE mg L^{-1}	MAE mg L^{-1}
66 (origin)	0.6727	0.7945 [0.6668, 0.8640]	20.62	13.83	0.6602	0.8335 [0.6929, 0.9110]	21	14.64
66 (origin) + 66 (virtual)	0.8437	0.8972 [0.8581, 0.9247]	13.5	7.86	0.8459	0.9182 [0.8704, 0.9522]	13.41	8.1
66 (origin) + 132 (virtual)	0.8912	0.9265 [0.9023, 0.9457]	10.64	5.99	0.8797	0.9411 [0.9124, 0.9624]	11.18	6.28
66 (origin) + 198 (virtual)	0.9182	0.9445 [0.9290, 0.9576]	9.18	4.77	0.9134	0.9586 [0.9394, 0.9748]	9.44	5.12
66 (origin) + 264 (virtual)	0.9249	0.9492 [0.9363, 0.9602]	8.62	4.42	0.9276	0.9630 [0.9477, 0.9763]	8.47	4.42
66 (origin) + 330 (virtual)	0.9357	0.9557 [0.9450, 0.9651]	7.88	3.94	0.9397	0.9700 [0.9568, 0.9809]	7.63	3.83
66 (origin) + 396 (virtual)	0.9436	0.9598 [0.9495, 0.9679]	7.26	3.66	0.9456	0.9726 [0.9609, 0.9831]	7.13	3.5
66 (origin) + 462 (virtual)	0.9505	0.9648 [0.9573, 0.9712]	6.83	3.35	0.9499	0.9763 [0.9666, 0.9849]	6.87	3.18
66 (origin) + 528 (virtual)	0.9524	0.9661 [0.9596, 0.9721]	6.49	3.17	0.9552	0.9773 [0.9683, 0.9847]	6.3	2.96
66 (origin) + 594 (virtual)	0.956	0.9680 [0.9622, 0.9738]	6.15	2.99	0.9575	0.9798 [0.9714, 0.9870]	6.04	2.78
66 (origin) + 660 (virtual)	0.9583	0.9701 [0.9646, 0.9752]	6.02	2.96	0.9622	0.9817 [0.9744, 0.9881]	5.73	2.56

Table S3. Performance evaluation of classical random forest and quantum feature enhanced random forest during the dry season under different numbers of virtual samples.

Sample size	RF				Quantum-enhanced RF			
	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE $mg\ L^{-1}$	MAE $mg\ L^{-1}$	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE $mg\ L^{-1}$	MAE $mg\ L^{-1}$
65 (origin)	0.2841	0.5151 [0.2884, 0.7276]	47.68	24.04	0.2883	0.6207 [0.3242, 0.8625]	47.54	25.26
65 (origin) + 65 (virtual)	0.5269	0.6603 [0.4760, 0.8160]	31.92	14.72	0.4726	0.7192 [0.4742, 0.8886]	33.71	16.44
65 (origin) + 130 (virtual)	0.6116	0.7255 [0.5790, 0.8380]	26.25	11.83	0.5712	0.7748 [0.5789, 0.9090]	27.58	12.98
65 (origin) + 195 (virtual)	0.609	0.7134 [0.5751, 0.8413]	22.52	8.71	0.5953	0.7915 [0.6036, 0.9267]	22.91	9.03
65 (origin) + 260 (virtual)	0.7392	0.8036 [0.7091, 0.8831]	19.7	8.02	0.7196	0.8538 [0.7269, 0.9424]	20.42	8.63
65 (origin) + 325 (virtual)	0.7331	0.8007 [0.7122, 0.8763]	18.3	7.27	0.7227	0.8548 [0.7330, 0.9406]	18.65	7.63
65 (origin) + 390 (virtual)	0.7873	0.8463 [0.7848, 0.9008]	17.31	6.9	0.7847	0.8911 [0.8036, 0.9523]	17.41	7.05
65 (origin) + 455 (virtual)	0.8251	0.8681 [0.8202, 0.9092]	15.81	6.43	0.8197	0.9073 [0.8363, 0.9583]	16.05	6.21
65 (origin) + 520 (virtual)	0.828	0.8696 [0.8222, 0.9098]	15.25	6.35	0.8292	0.9024 [0.8352, 0.9520]	15.2	6.27
65 (origin) + 585 (virtual)	0.8542	0.8862 [0.8431, 0.9221]	14.36	6.1	0.8422	0.9124 [0.8576, 0.9538]	14.94	5.89
65 (origin) + 650 (virtual)	0.8478	0.8811 [0.8408, 0.9178]	13.99	5.87	0.844	0.9203 [0.8648, 0.9637]	14.17	5.44

Table S4. Performance evaluation of classical random forest and quantum feature enhanced random forest during the wet season under different numbers of virtual samples.

Sample size	RF				Quantum-enhanced RF			
	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE $mg\ L^{-1}$	MAE $mg\ L^{-1}$	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE $mg\ L^{-1}$	MAE $mg\ L^{-1}$
50 (origin)	0.7331	0.8219 [0.6715, 0.8926]	12.19	7.95	0.7166	0.8552 [0.7192, 0.9305]	12.56	8.61
50 (origin) + 50 (virtual)	0.8402	0.8984 [0.8323, 0.9371]	8.13	4.44	0.829	0.9152 [0.8405, 0.9627]	8.41	4.88

50 (origin) + 100 (virtual)	0.9209	0.9486 [0.9262, 0.9657]	6.14	3.04	0.9168	0.9583 [0.9309, 0.9794]	6.3	3.38
50 (origin) + 150 (virtual)	0.9409	0.9612 [0.9461, 0.9728]	5.25	2.5	0.9362	0.9686 [0.9503, 0.9837]	5.45	2.81
50 (origin) + 200 (virtual)	0.9575	0.9715 [0.9607, 0.9804]	4.55	1.99	0.9566	0.9767 [0.9640, 0.9878]	4.6	2.05
50 (origin) + 250 (virtual)	0.9615	0.9741 [0.9652, 0.9810]	4.14	1.81	0.9594	0.9802 [0.9693, 0.9892]	4.25	1.88
50 (origin) + 300 (virtual)	0.9691	0.9787 [0.9722, 0.9837]	3.78	1.53	0.9672	0.9845 [0.9763, 0.9916]	3.9	1.61
50 (origin) + 350 (virtual)	0.9739	0.9814 [0.9759, 0.9857]	3.45	1.44	0.973	0.9867 [0.9797, 0.9929]	3.51	1.44
50 (origin) + 400 (virtual)	0.9759	0.9827 [0.9769, 0.9868]	3.28	1.35	0.9744	0.9878 [0.9813, 0.9935]	3.39	1.32
50 (origin) + 450 (virtual)	0.9762	0.9840 [0.9788, 0.9880]	3.25	1.27	0.9756	0.9889 [0.9829, 0.9940]	3.293	1.23
50 (origin) + 500 (virtual)	0.9774	0.9836 [0.9787, 0.9875]	3.03	1.18	0.9739	0.9884 [0.9823, 0.9936]	3.26	1.2

Table S5. Performance evaluation of classical random forest and quantum feature-enhanced random forest under different virtual sample numbers in the normal season based on AlphaEarth Foundation as the input variable.

Sample size	RF				Quantum-enhanced RF			
	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE mg L^{-1}	MAE mg L^{-1}	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE mg L^{-1}	MAE mg L^{-1}
66 (origin)	0.1672	0.5508 [0.3178, 0.6978]	32.89	25.4	0.1193	0.5320 [0.3098, 0.6799]	33.82	26.41
66 (origin) + 66 (virtual)	0.3772	0.6457 [0.4886, 0.7547]	25.24	17.53	0.3315	0.6344 [0.4668, 0.7474]	26.15	18.39
66 (origin) + 132 (virtual)	0.5832	0.7662 [0.6731, 0.8300]	21.21	13.63	0.5626	0.7500 [0.6564, 0.8168]	21.73	14.54
66 (origin) + 198 (virtual)	0.604	0.7692 [0.6931, 0.8300]	20	11.9	0.5829	0.7616 [0.6808, 0.8244]	20.53	12.78
66 (origin) + 264 (virtual)	0.7184	0.8363 [0.7874, 0.8794]	17.33	9.31	0.7118	0.8302 [0.7780, 0.8753]	17.53	9.95

66 (origin) + 330 (virtual)	0.7609	0.8557 [0.8069, 0.8957]	15.46	8.08	0.7446	0.8497 [0.7986, 0.8897]	15.98	8.82
66 (origin) + 396 (virtual)	0.7825	0.8673 [0.8279, 0.8991]	14.53	7.29	0.7689	0.8639 [0.8256, 0.8965]	14.97	7.96
66 (origin) + 462 (virtual)	0.8232	0.8927 [0.8584, 0.9223]	12.99	6.04	0.8176	0.8906 [0.8557, 0.9206]	13.19	6.42
66 (origin) + 528 (virtual)	0.8327	0.8895 [0.8525, 0.9184]	11.84	5.26	0.8045	0.8797 [0.8366, 0.9107]	12.8	6.29
66 (origin) + 594 (virtual)	0.8524	0.9038 [0.8699, 0.9323]	10.87	4.63	0.8326	0.8925 [0.8553, 0.9225]	11.58	5.57
66 (origin) + 660 (virtual)	0.8595	0.9093 [0.8792, 0.9326]	10.73	4.46	0.8438	0.8991 [0.8690, 0.9235]	11.32	5.42

Table S6. Performance evaluation of classical random forest and quantum feature-enhanced random forest under different virtual sample numbers in the dry season based on AlphaEarth Foundation as the input variable.

Sample size	RF				Quantum-enhanced RF			
	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE mg L^{-1}	MAE mg L^{-1}	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE mg L^{-1}	MAE mg L^{-1}
65 (origin)	-0.0737	0.3844 [0.0365, 0.6712]	58.96	35.78	-0.0613	0.3780 [0.0313, 0.6531]	58.62	35.73
65 (origin) + 65 (virtual)	0.0391	0.4235 [0.1666, 0.6633]	42.02	22.64	0.0021	0.4214 [0.1673, 0.6550]	42.89	23.3
65 (origin) + 130 (virtual)	0.3521	0.6014 [0.4262, 0.7627]	37.36	19.52	0.3096	0.5965 [0.4144, 0.7529]	38.56	21.35
65 (origin) + 195 (virtual)	0.4229	0.6609 [0.5062, 0.7877]	34.52	16.36	0.4029	0.6582 [0.5072, 0.7858]	35.11	17.38
65 (origin) + 260 (virtual)	0.5655	0.7400 [0.6216, 0.8389]	30.59	13.51	0.5567	0.7339 [0.6154, 0.8295]	30.9	14.89
65 (origin) + 325 (virtual)	0.5979	0.7517 [0.6342, 0.8568]	27.1	11.1	0.5743	0.7332 [0.6056, 0.8407]	28.71	12.6
65 (origin) + 390 (virtual)	0.6345	0.7786 [0.6882, 0.8676]	26.57	10.67	0.6273	0.7740 [0.6829, 0.8630]	26.83	11.2
65 (origin) + 455 (virtual)	0.6705	0.7994 [0.7125, 0.8790]	23.57	8.59	0.6607	0.7992 [0.7079, 0.8790]	23.91	9.64

65 (origin) + 520 (virtual)	0.5911	0.7520 [0.6398, 0.8608]	22.55	8.11	0.5892	0.7474 [0.6317, 0.8520]	22.6	8.97
65 (origin) + 585 (virtual)	0.641	0.7661 [0.6385, 0.8638]	21.05	7.42	0.6306	0.7528 [0.6359, 0.8450]	21.58	8.32
65 (origin) + 650 (virtual)	0.6741	0.7968 [0.7123, 0.8810]	19.86	6.41	0.6762	0.7945 [0.7078, 0.8787]	19.69	7.03

Table S7. Performance evaluation of classical random forest and quantum feature-enhanced random forest under different virtual sample numbers in the wet season based on AlphaEarth Foundation as the input variable.

Sample size	RF				Quantum-enhanced RF			
	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE	MAE	R^2 LOOCV	R^2 [95% CI] Bootstrap	RMSE	MAE
50 (origin)	-0.2089	0.3864 [0.1382, 0.5823]	25.16	19.8	-0.1938	0.3865 [0.1271, 0.5816]	25	19.83
50 (origin) + 50 (virtual)	0.3662	0.6567 [0.5386, 0.7519]	15.79	11.07	0.3643	0.6410 [0.5166, 0.7382]	15.82	11.33
50 (origin) + 100 (virtual)	0.5157	0.7338 [0.6583, 0.8014]	13.47	8.7	0.4945	0.7245 [0.6489, 0.7934]	13.77	9.32
50 (origin) + 150 (virtual)	0.5677	0.7646 [0.7022, 0.8213]	12.44	7.68	0.5577	0.7555 [0.6958, 0.8107]	12.62	8.19
50 (origin) + 200 (virtual)	0.6162	0.7788 [0.7216, 0.8323]	11.35	6.39	0.5965	0.7709 [0.7157, 0.8222]	11.64	6.98
50 (origin) + 250 (virtual)	0.6388	0.7966 [0.7466, 0.8425]	11.1	6.07	0.6142	0.7818 [0.7337, 0.8236]	11.47	6.84
50 (origin) + 300 (virtual)	0.6847	0.8208 [0.7769, 0.8626]	10.44	5.33	0.6805	0.8128 [0.7701, 0.8510]	10.51	5.78
50 (origin) + 350 (virtual)	0.7175	0.8335 [0.7898, 0.8691]	9.69	4.86	0.704	0.8274 [0.7850, 0.8631]	9.92	5.34
50 (origin) + 400 (virtual)	0.7478	0.8531 [0.8134, 0.8871]	8.96	4.31	0.7349	0.8436 [0.8067, 0.8771]	9.18	4.77
50 (origin) + 450 (virtual)	0.78	0.8677 [0.8306, 0.8984]	8.39	3.93	0.7631	0.8594 [0.8237, 0.8892]	8.71	4.39
50 (origin) + 500 (virtual)	0.7841	0.8731 [0.8397, 0.9024]	8.27	3.8	0.781	0.8650 [0.8320, 0.8920]	8.33	4.15