Supplementary materials of

Coupling compositional data analysis (CoDA) with hierarchical cluster analysis (HCA) for preliminary understanding of the dynamics of a complex water distribution system: the Naples (south Italy) case study

Pooria Ebrahimi[a], Stefano Albanese[a*], Leopoldo Esposito[b], Daniela Zuzolo[c], Domenico Cicchella[c]

[a] *Department of Earth, Environmental and Resources Sciences, University of Naples Federico II, Napoli 80126, Italy*
[b] *Acqua Bene Comune Napoli Azienda Speciale, Napoli 80147, Italy*
[c] *Department of Science and Technology, University of Sannio, Benevento 82100, Italy*

**Corresponding author: stefano.albanese@unina.it*

Table S1 Coordinates and address of the stations for regular tap water sampling in Naples.

Table S2 Composition of the input water in the water distribution system of Naples in 2017.

Fig. S1 Distribution of the measured EC, NO_3^- and pH values in the input waters for each season of 2017. The numbers above boxplots show the number of measurements in the corresponding seasons. (a-c) I1: ACO; (d-f) I2: AIR wells; (g-i) I3: Cancello well field; (j-l) I4: Campano Aqueduct and Ponte Tavano wells; (m-o) I5: Serino Aqueduct.

Fig. S2 Seasonal compositions of the input waters in the water distribution system of Naples in 2017: (a) I1: ACO; (b) I2: AIR wells; (c) I3: Cancello well field; (d) I4: Campano Aqueduct and Ponte Tavano wells; and (e) I5: Serino Aqueduct.

Fig. S3 Seasonal composition of the tap waters originated from: (a) R1: Capodimonte; (b) R2: S. Sebastiano; (c) R3: Scudillo; (d) R4: S. Giacomo; (e) R5: Pianura; (f) R6: Cangiani; (g) R7: Camaldoli; (h) S1: ACO; (i) S2: Cancello well field, Campano Aqueduct and Ponte Tavano wells; and (j) S3: Serino Aqueduct.

Fig. S4 Variation of (a) As, (b) Ni, (c) V, (d) Fe, (e) Mn, (f) Al, (g) ClO_2^-, (h) Cl_2, (i) Na^+, (j) K^+, (k) Ca^{2+}, (l) Mg^{2+}, (m) HCO_3^-, (n) Cl^-, (o) SO_4^{2-}, (p) NO_3^-, (q) F^-, (r) EC, (s) turbidity, (t) pH and (u) temperature in the collected tap water samples. The gray circles indicate the monthly measurements and the black squares show the annual average values. The black dashed lines represents drinking water action limits (See Tables 2 and S1). R1: Capodimonte; R2: S. Sebastiano; R3: Scudillo; R4: S. Giacomo; R5: Pianura; R6: Cangiani; R7: Camaldoli; S1: ACO; S2: Cancello well field, Campano Aqueduct and Ponte Tavano wells; S3: Serino Aqueduct; and UN: Not defined in the water supply scheme.

Fig. S5 Spatial representation of the results of seasonal hierarchical cluster analysis: (a) winter; (b) spring; (c) summer; and (d) fall. R1: Capodimonte; R2: S. Sebastiano; R3: Scudillo; S1: ACO; and S3: Serino Aqueduct.

Fig. S6 The tap water composition before and after (without and with asterisk, respectively) implementing the HCA result: (a) R1: Capodimonte; (b) R3: Scudillo; (c) S2: Cancello well field, Campano Aqueduct and Ponte Tavano wells, I3: Cancello well field, I4: Campano Aqueduct and Ponte Tavano wells, and I5: Serino Aqueduct; and (d) S3: Serino Aqueduct.

Text S1 HCA performance

Table S1 Coordinates and address of the stations for regular tap water sampling in Naples.

Station	Latitude	Longitude	Address
D1	40.83293	14.22750	Riviera di Chiaia
D2	40.83725	14.23346	Piazza Amedeo
D3	40.80584	14.20356	Via Posillipo
D4	40.83105	14.21260	Via Manzoni
D5	40.82864	14.24821	Piazzetta Marinari
D6	40.83786	14.24191	Piazza Mondragone
D7	40.82834	14.19706	Via Benedetto Cariteo
D8	40.81558	14.18782	Via Cavalleggeri d'Aosta
D9	40.81598	14.16891	Via S. Ferrara
D10	40.81671	14.16007	Via di Pozzuoli
D11	40.84262	14.19819	Via Adriano
D12	40.84828	14.19665	Via Scherillo
D13	40.85508	14.18035	Via Padula
D14	40.86217	14.17434	Corso Duca d'Aosta
D15	40.84716	14.22119	Via Caldieri
D16	40.85015	14.22429	Via Simone Martini
D17	40.84272	14.22162	Via Santa Maria Della Libera
D18	40.86547	14.22863	Via del Serbatoio
D19	40.84402	14.23613	Via Ligorio Pirro
D20	40.88852	14.24409	Via V. Veneto
D21	40.88952	14.22445	Via E. Scaglione
D22	40.89018	14.21661	Corso Chiaiano
D23	40.86793	14.21358	Via L. Bianchi
D24	40.86862	14.19166	Via Guantai ad Orsolona
D25	40.90044	14.24549	Via A. Ghisleri
D26	40.86703	14.26864	Via Arenaccia
D27	40.86916	14.25448	Via Ponti Rossi
D28	40.87105	14.24460	Viale Colli Aminei
D29	40.86182	14.24565	Vicoletto S. Vincenzo alla Sanità
D30	40.85501	14.25383	Piazza Cavour
D31	40.89304	14.25370	Via Lombardia
D32	40.88509	14.23899	Via G. Manfredi ang. Via Lieti a Capodimonte
D33	40.87699	14.27734	Varco ingresso dogana Aeroporto - Viale U. Maddalena
D34	40.89060	14.28059	Via G. Pascale
D35	40.87987	14.30715	Via Icaro
D36	40.84251	14.24419	Corso V. Emanuele
D37	40.84932	14.24976	Piazza Dante
D38	40.85388	14.24652	Via S. Rosa
D39	40.85024	14.26379	Piazza Calenda
D40	40.84766	14.26880	Corso Garibaldi
D41	40.84137	14.25368	Via Flavio Gioia
D42	40.84685	14.25189	Piazza del Gesu
D43	40.82874	14.31460	Via Bernardino Martirano
D44	40.86975	14.33157	Via Madonnelle
D45	40.85177	14.33326	C.so Ponticelli
D46	40.84571	14.33017	Via Mastellone
D47	40.84610	14.31420	Via Volpicella
D48	40.85672	14.26142	Via C. Rosaroll
D49	40.85710	14.27124	C.ta Ponte di Casanova
D51	40.84670	14.28033	Via Brin
D52	40.86529	14.29121	Largo Tarantini

Table S2 Composition of the input water in the water distribution system of Naples in 2017.

Input	Seasonal/ Annual	Ca^{2+}	Mg^{2+}	Na^{+}	SO_4^{2-}	Cl^{-}	F^{-}	NO_3^{-}	EC	pH
		mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	µS/cm	-
I1 [a]	Winter	121.6	28.6	8.0	15.0	11.0	0.2	3.9	724.3	7.1
	Spring	124.3	31.0	10.0	18.0	16.0	0.2	4.2	722.9	7.2
	Summer	116.8	28.5	8.0	18.0	12.0	0.2	4.1	736.2	7.2
	Fall	123.8	30.2	8.0	22.0	13.0	0.2	3.8	726.6	7.1
	Annual	121.6	29.6	8.5	18.3	13.0	0.2	4.0	727.5	7.2
I2 [b]	Winter	131.8	27.0	-	-	86.9	1.8	106.3	1172.0	7.2
	Spring	136.8	28.1	80.0	168.0	89.2	1.7	111.1	1165.8	7.3
	Summer	159.8	32.0	-	-	124.9	1.5	84.6	1265.1	7.1
	Fall	153.8	33.5	84.0	152.0	105.1	2.0	104.8	1170.2	7.2
	Annual	145.3	30.2	82.0	160.0	102.0	1.8	101.8	1193.2	7.2
I3 [c]	Winter	165.2	40.4	87.0	46.0	172.0	0.6	9.5	1410.7	6.9
	Spring	185.2	49.4	97.0	52.0	197.0	0.6	9.1	1419.9	7.0
	Summer	190.9	48.2	102.0	51.0	201.0	0.4	9.1	1489.5	6.9
	Fall	172.9	43.1	88.0	43.0	164.0	0.6	9.3	1343.0	6.9
	Annual	178.6	45.3	93.5	48.0	183.5	0.5	9.2	1415.7	6.9
I4 [d]	Winter	188.0	51.8	102.0	59.0	200.0	0.8	13.0	1596.2	7.0
	Spring	-	-	-	-	-	-	13.0	1607.7	7.0
	Summer	-	-	-	-	-	-	12.7	1560.3	7.1
	Fall	202.5	55.6	100.0	58.0	217.0	0.7	13.2	1615.7	6.9
	Annual	195.3	53.7	101.0	58.5	208.5	0.7	13.0	1596.6	7.0
I5 [e]	Winter	54.1	9.6	7.0	9.0	9.0	0.2	4.7	351.2	7.7
	Spring	-	-	-	-	-	-	4.2	347.4	7.8
	Summer	55.0	8.9	7.0	5.0	10.0	0.2	4.0	335.7	7.9
	Fall	-	-	-	-	-	-	4.2	339.8	7.6
	Annual	54.6	9.3	7.0	7.0	9.5	0.2	4.3	345.2	7.8

[a] ACO; [b] AIR wells; [c] Cancello well field; [d] Campano Aqueduct and Ponte Tavano wells; [e] Serino Aqueduct. EC, NO_3^- and pH were regularly measured (see Fig. S1 for further details). For the rest of chemicals, results of six measurements in AIR wells and one measurement in the others are available in each season.

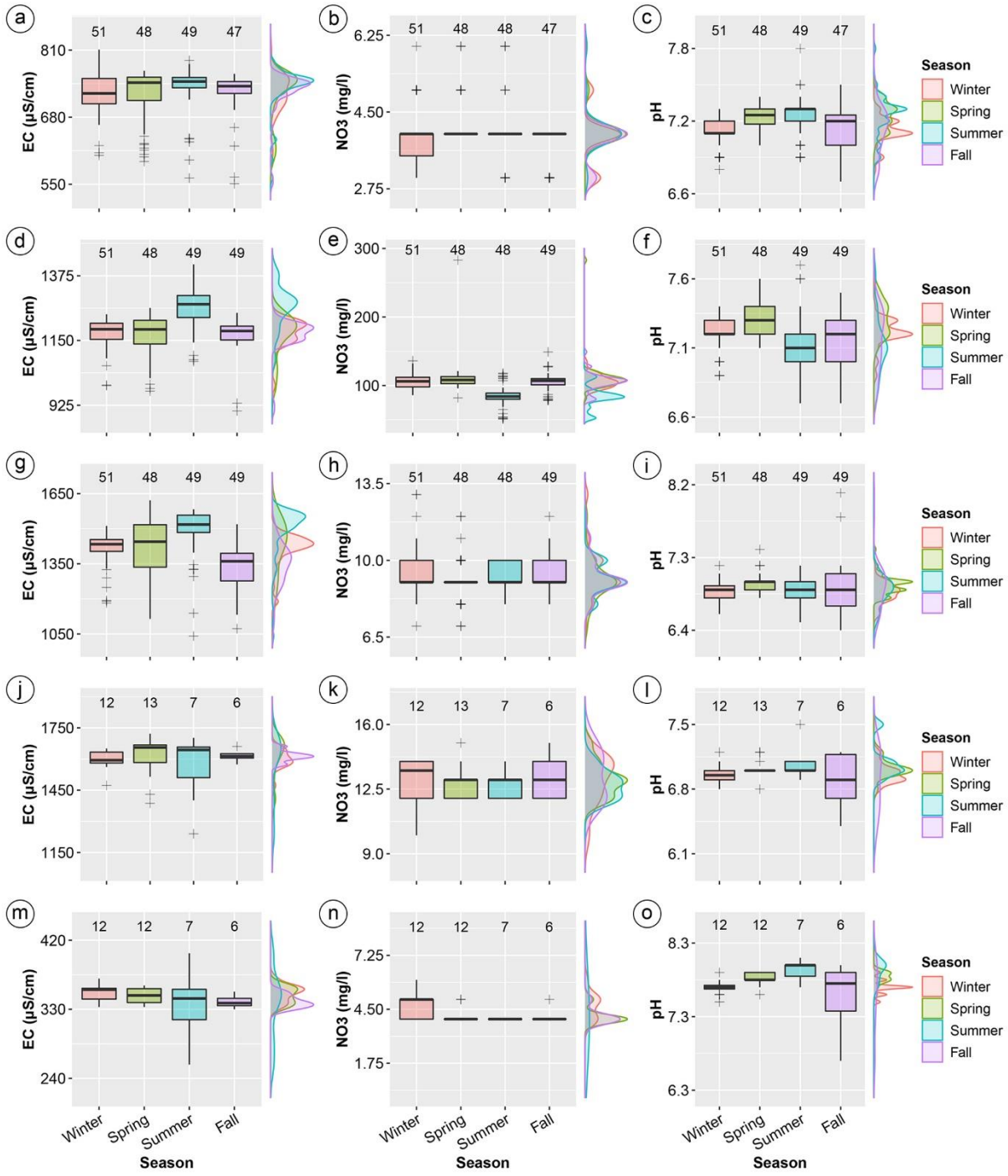


Fig. S1 Distribution of the measured EC, NO_3^- and pH values in the input waters for each season of 2017. The numbers above boxplots show the number of measurements in the corresponding seasons. (a-c) I1: ACO; (d-f) I2: AIR wells; (g-i) I3: Cancello well field; (j-l) I4: Campano Aqueduct and Ponte Tavano wells; (m-o) I5: Serino Aqueduct.

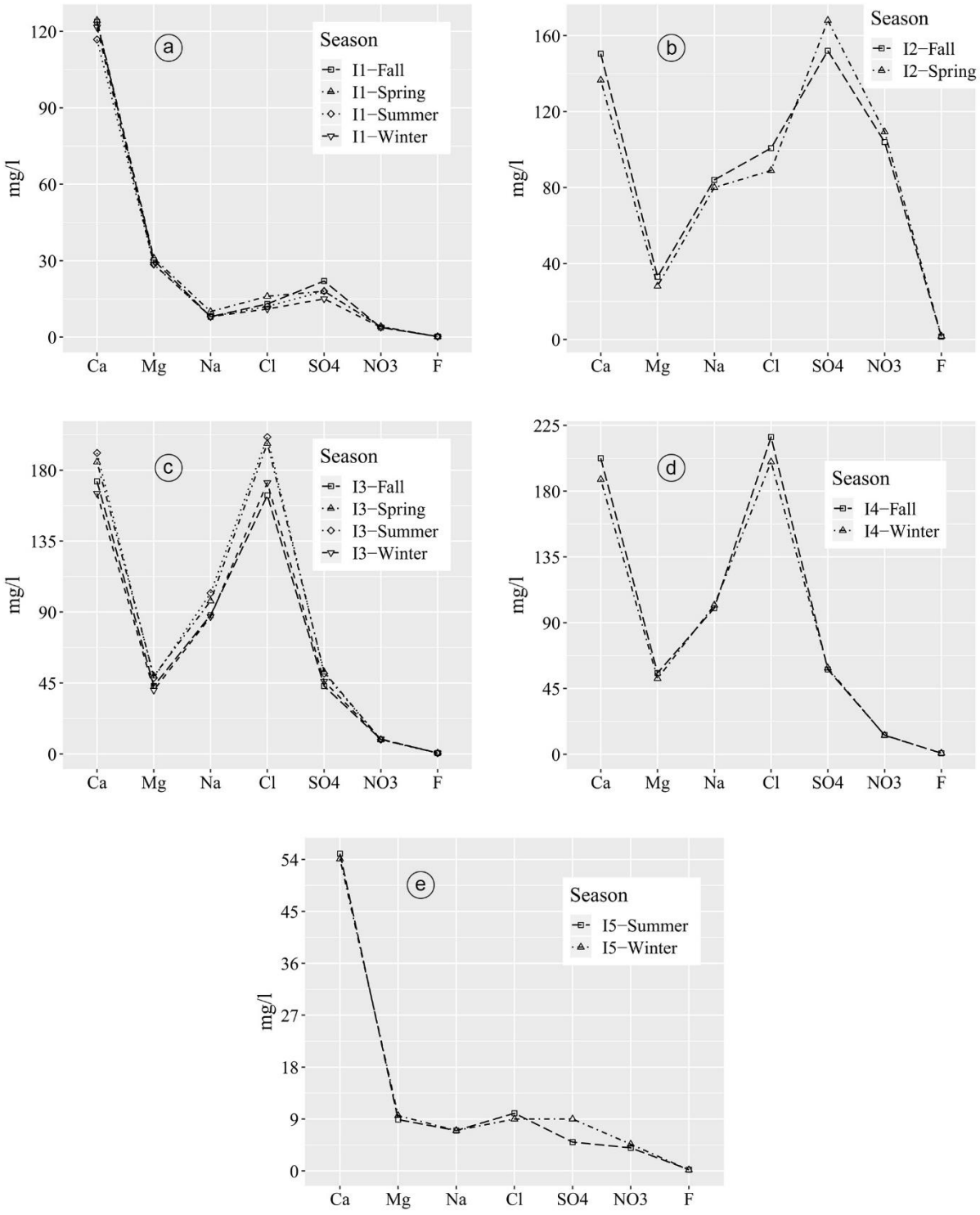


Fig. S2 Seasonal compositions of the input waters in the water distribution system of Naples in 2017: (a) I1: ACO; (b) I2: AIR wells; (c) I3: Cancello well field; (d) I4: Campano Aqueduct and Ponte Tavano wells; and (e) I5: Serino Aqueduct.

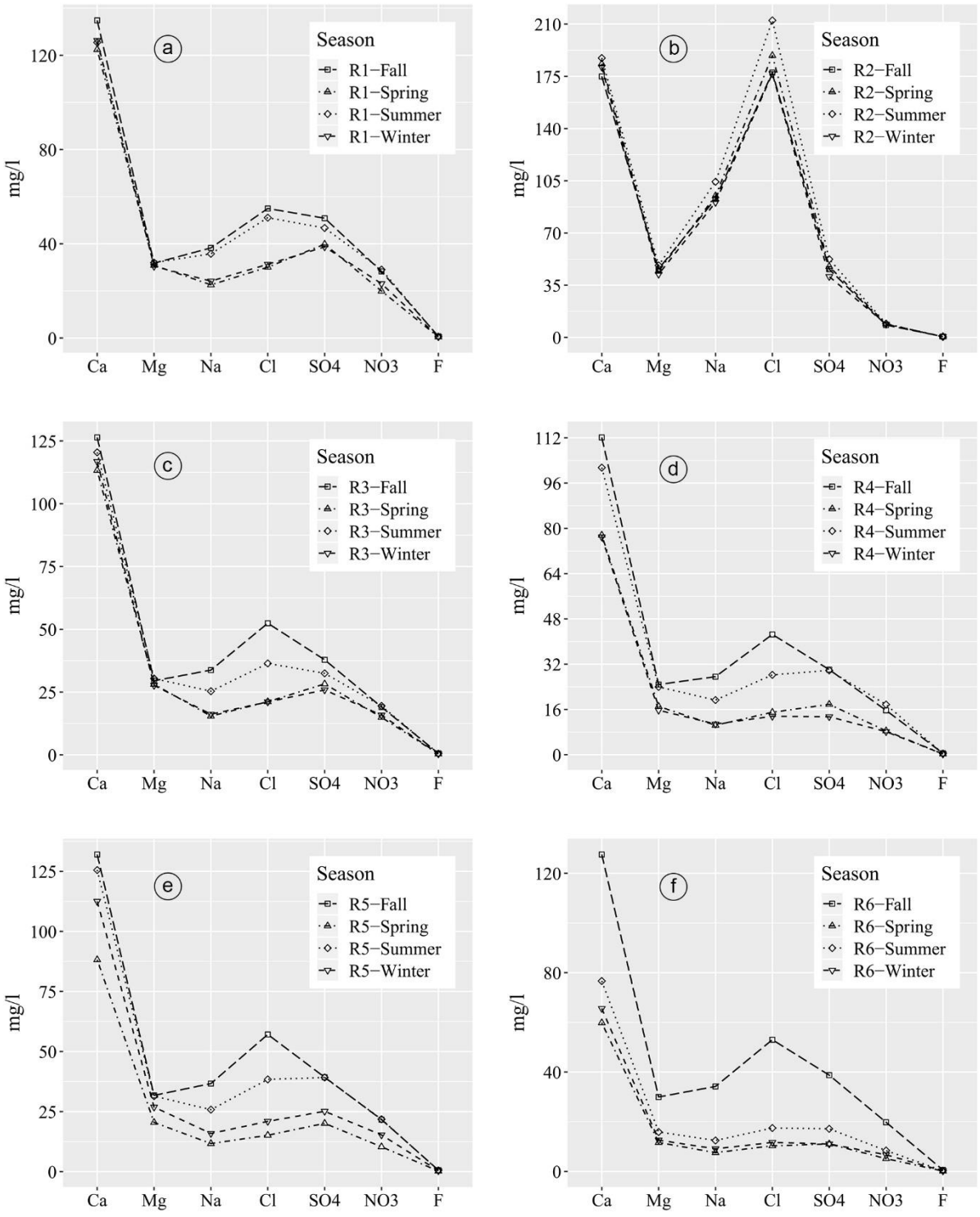


Fig. S3 Seasonal composition of the tap waters originated from: (a) R1: Capodimonte; (b) R2: S. Sebastiano; (c) R3: Scudillo; (d) R4: S. Giacomo; (e) R5: Pianura; (f) R6: Cangiani; (g) R7: Camaldoli; (h) S1: ACO; (i) S2: Cancello well field, Campano Aqueduct and Ponte Tavano wells; and (j) S3: Serino Aqueduct.

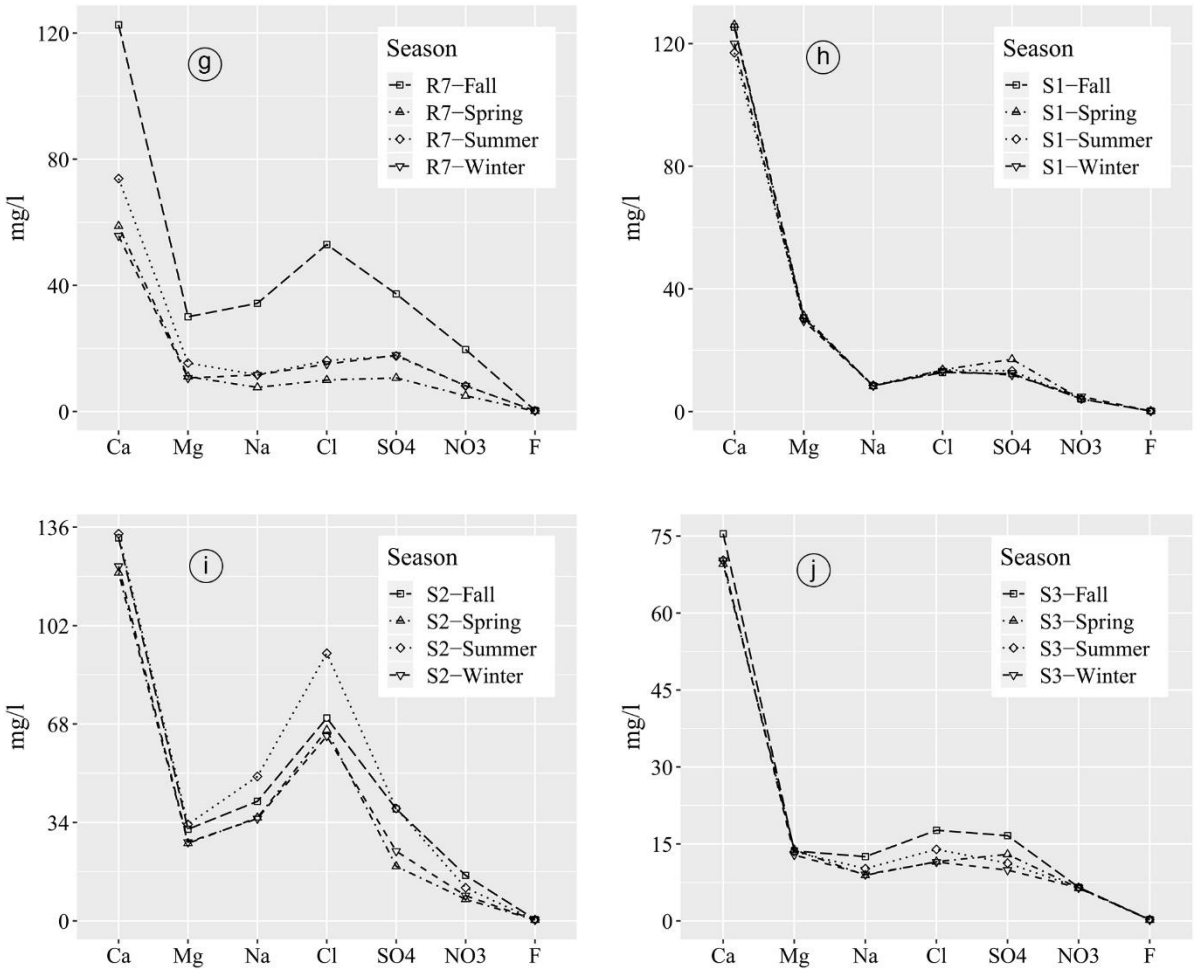


Fig. S3 (continued)

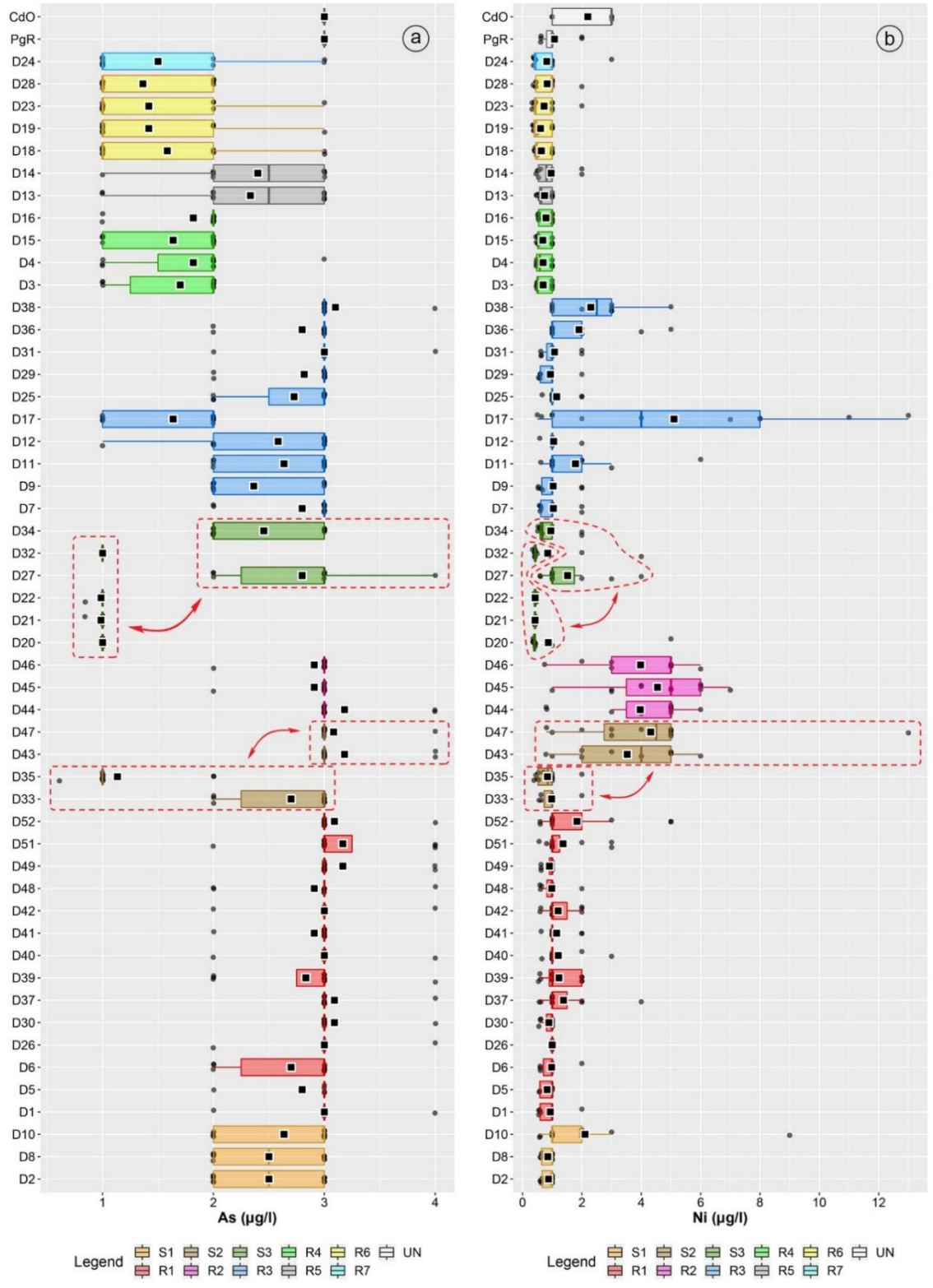


Fig. S4 Variation of (a) As, (b) Ni, (c) V, (d) Fe, (e) Mn, (f) Al, (g) ClO_2^-, (h) Cl_2, (i) Na^+, (j) K^+, (k) Ca^{2+}, (l) Mg^{2+}, (m) HCO_3^-, (n) Cl^-, (o) SO_4^{2-}, (p) NO_3^-, (q) F^-, (r) EC, (s) turbidity, (t) pH and (u) temperature in the collected tap water samples. The gray circles indicate the monthly measurements and the black squares show the annual average values. The black dashed lines represents drinking water action limits (See Tables 2 and S1). R1: Capodimonte; R2: S. Sebastiano; R3: Scudillo; R4: S. Giacomo; R5: Pianura; R6: Cangiani; R7: Camaldoli; S1: ACO; S2: Cancello well field, Campano Aqueduct and Ponte Tavano wells; S3: Serino Aqueduct; and UN: Not defined in the water supply scheme.

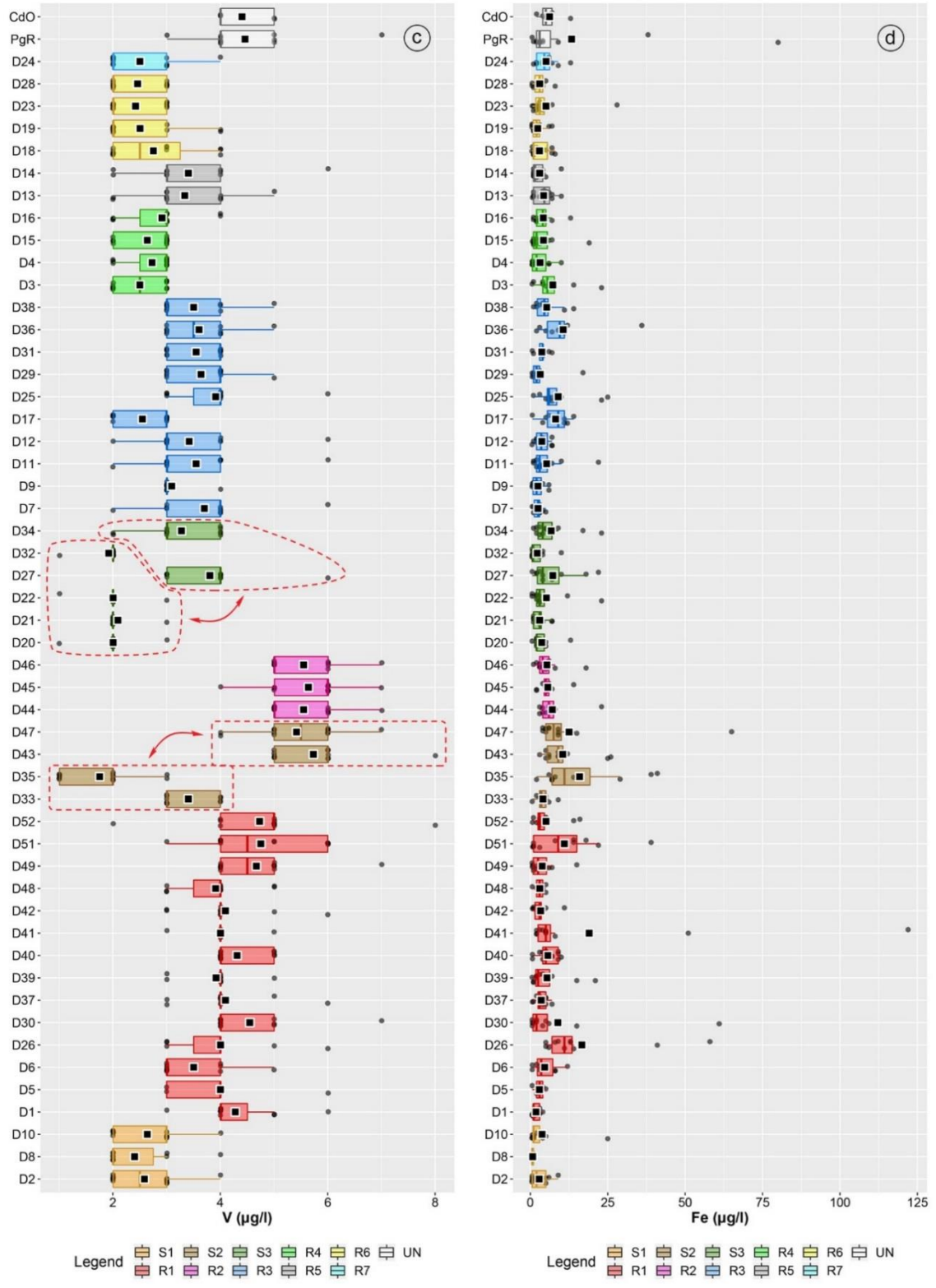


Fig. S4 (continued)

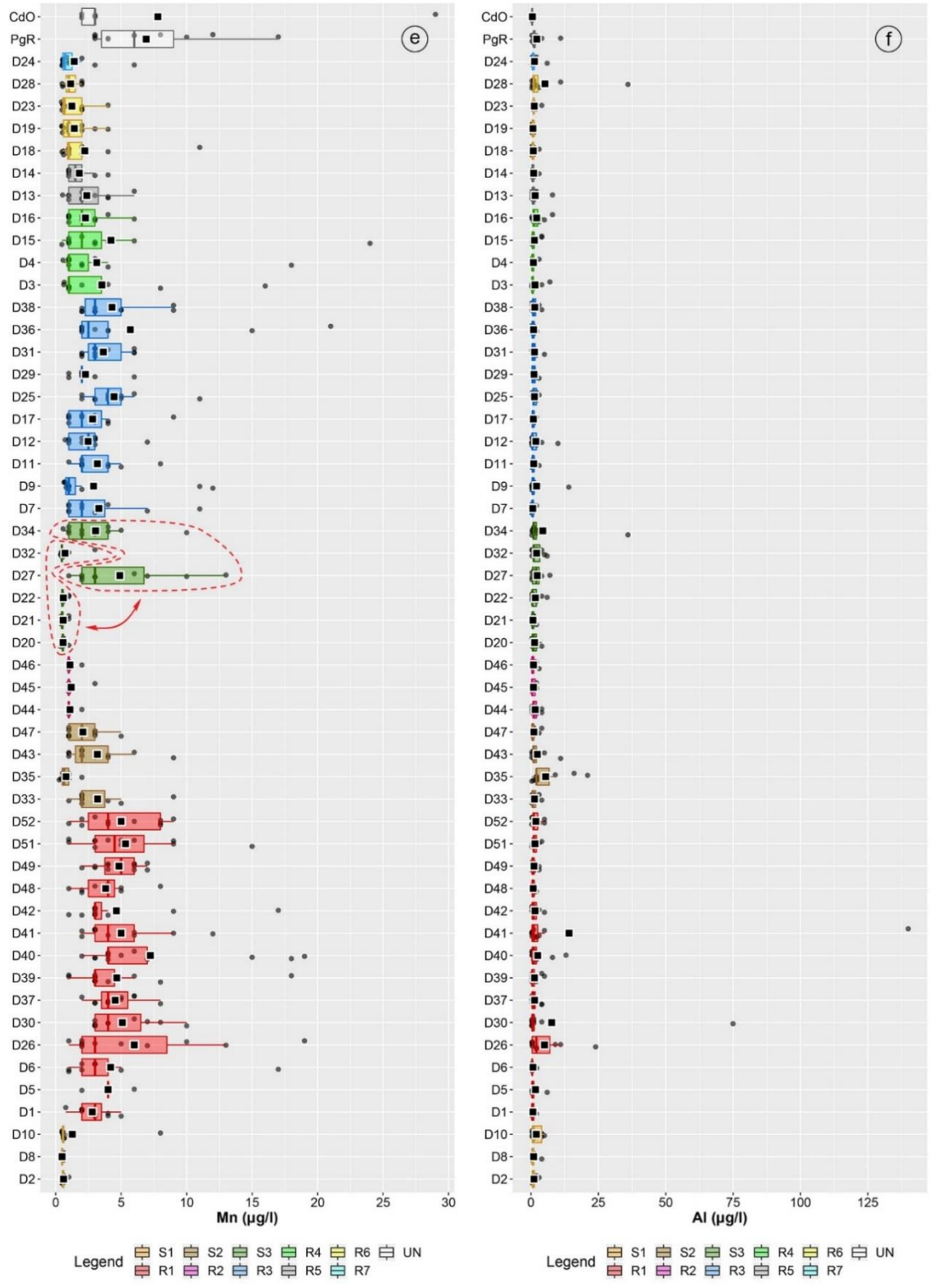


Fig. S4 (continued)

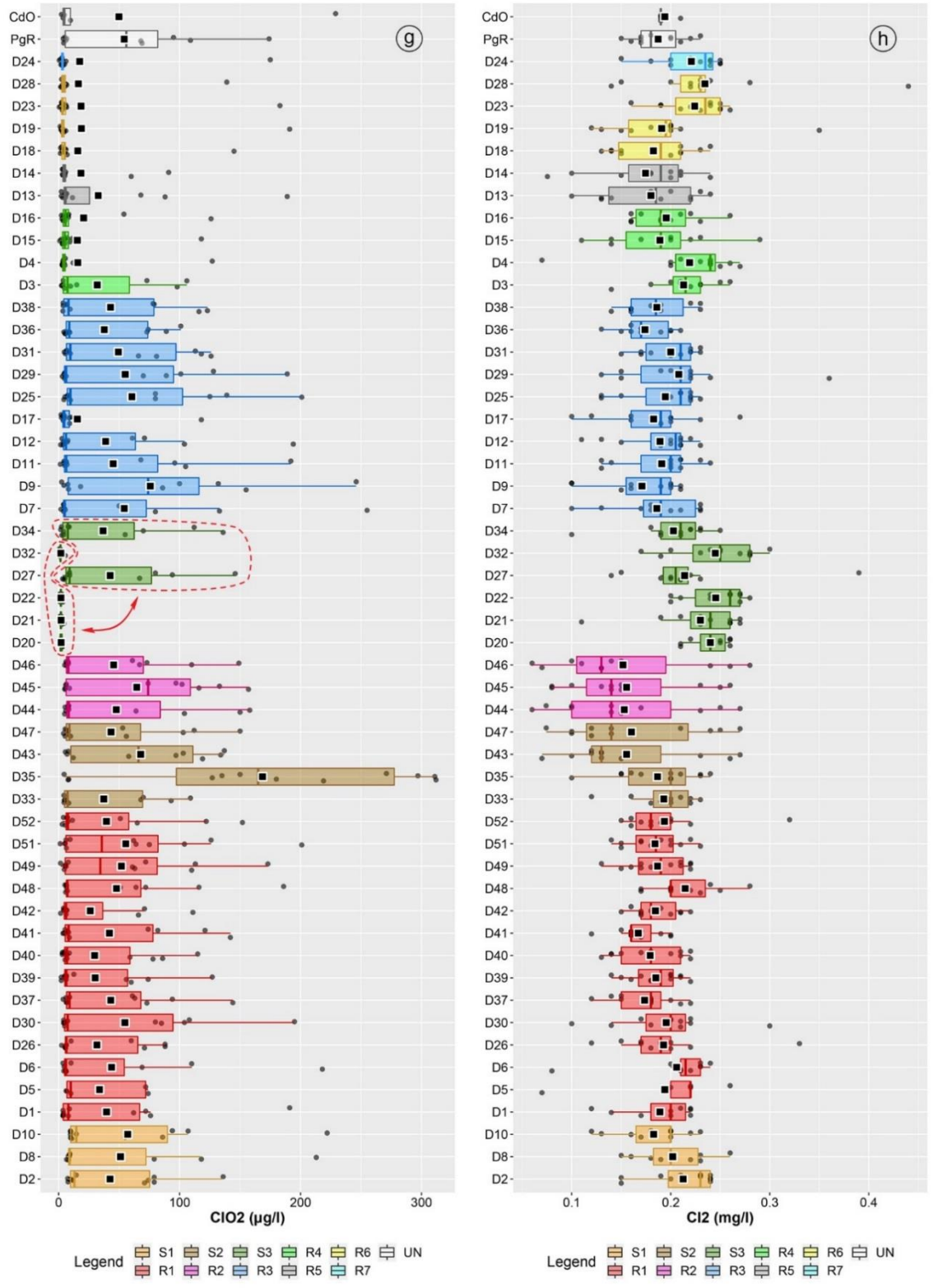


Fig. S4 (continued)

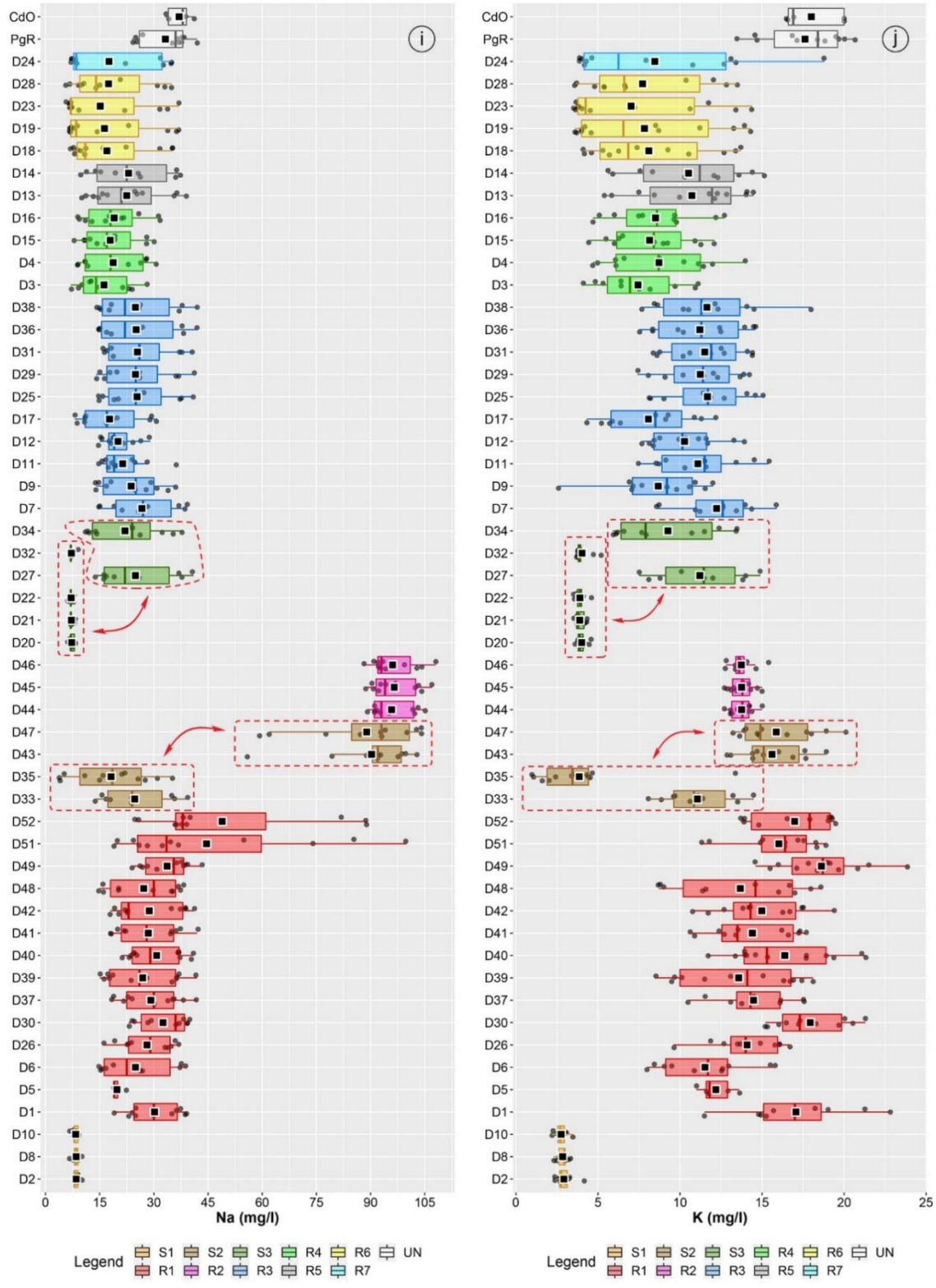


Fig. S4 (continued)

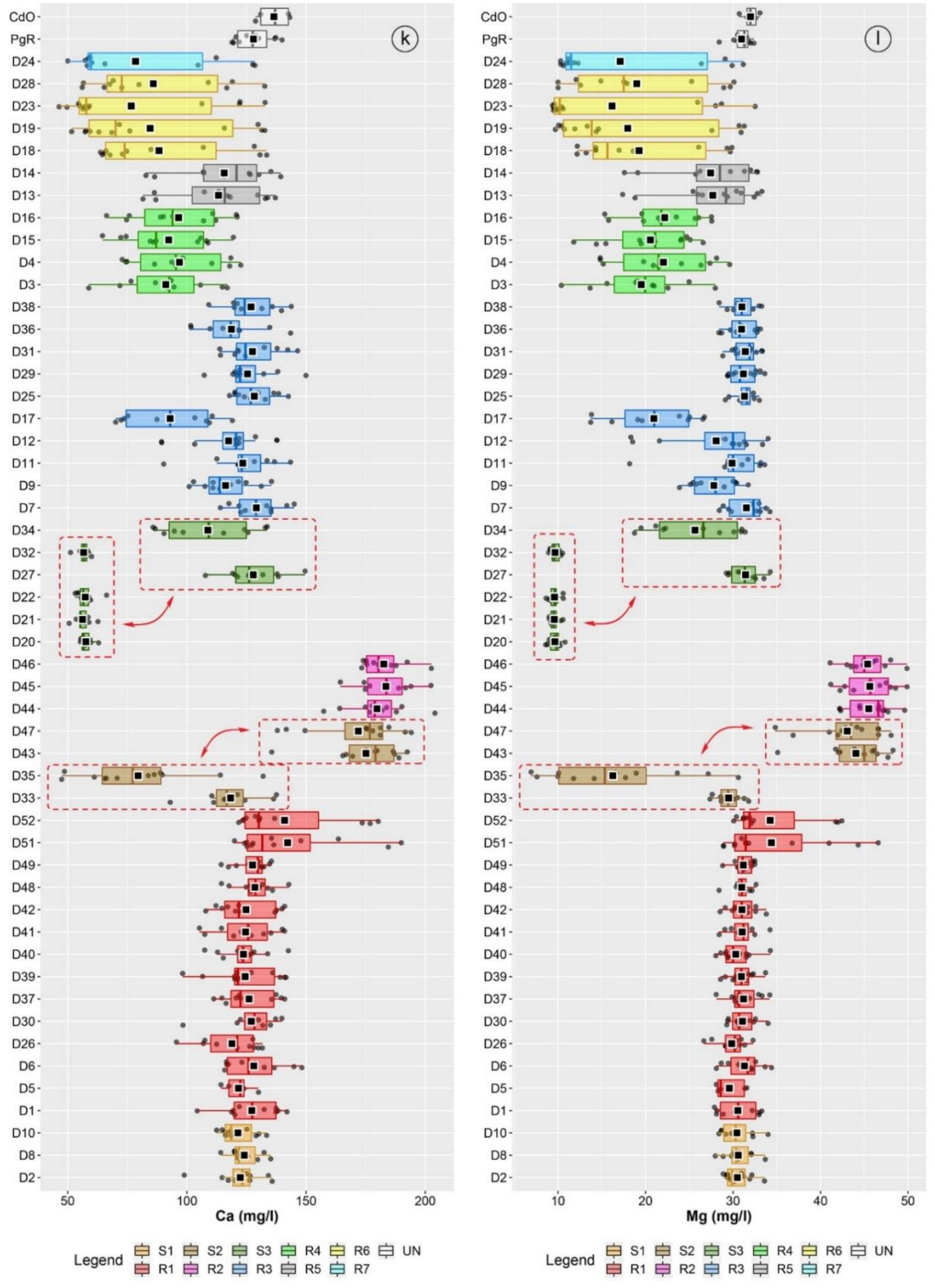


Fig. S4 (continued)

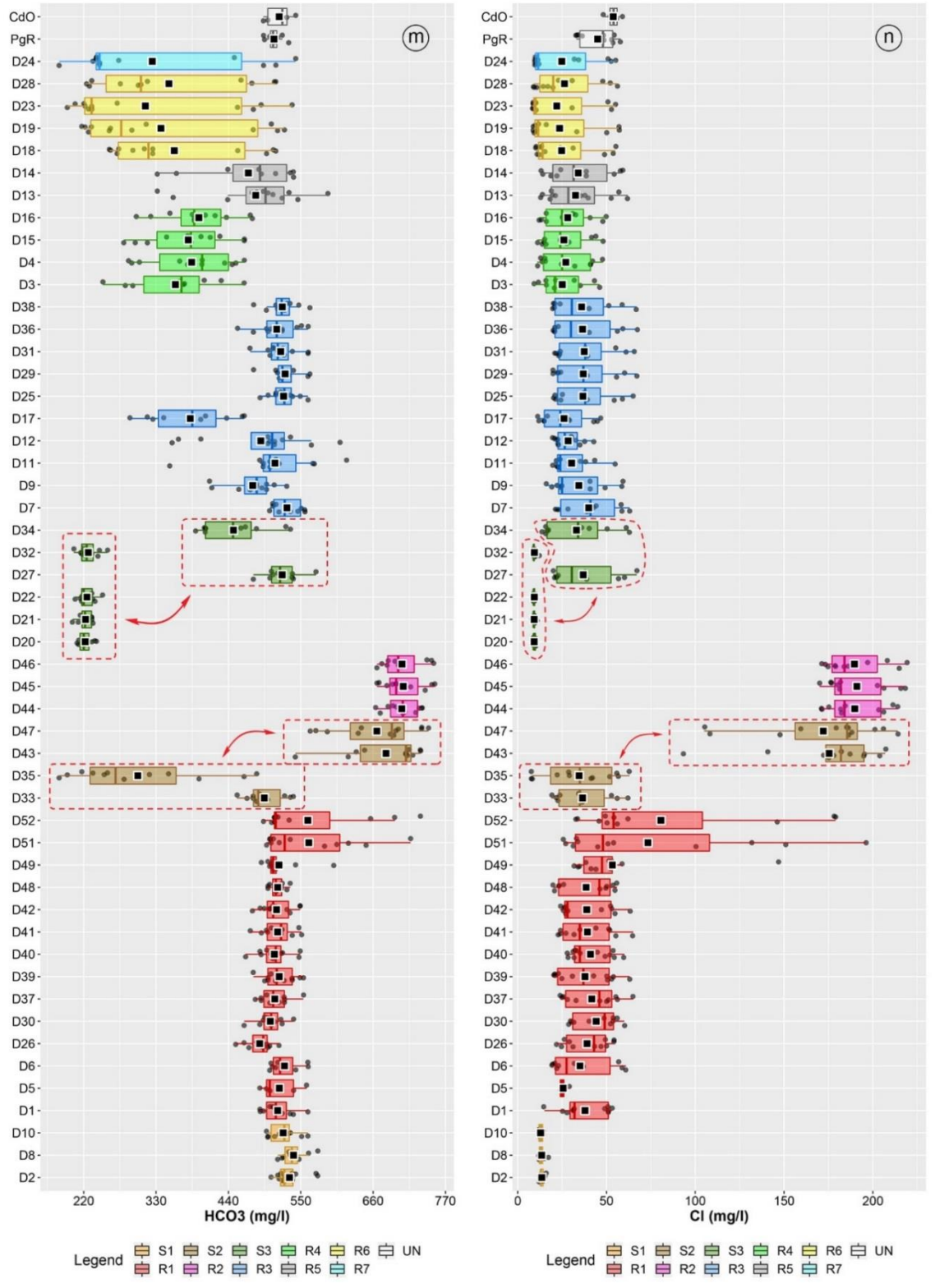


Fig. S4 (continued)

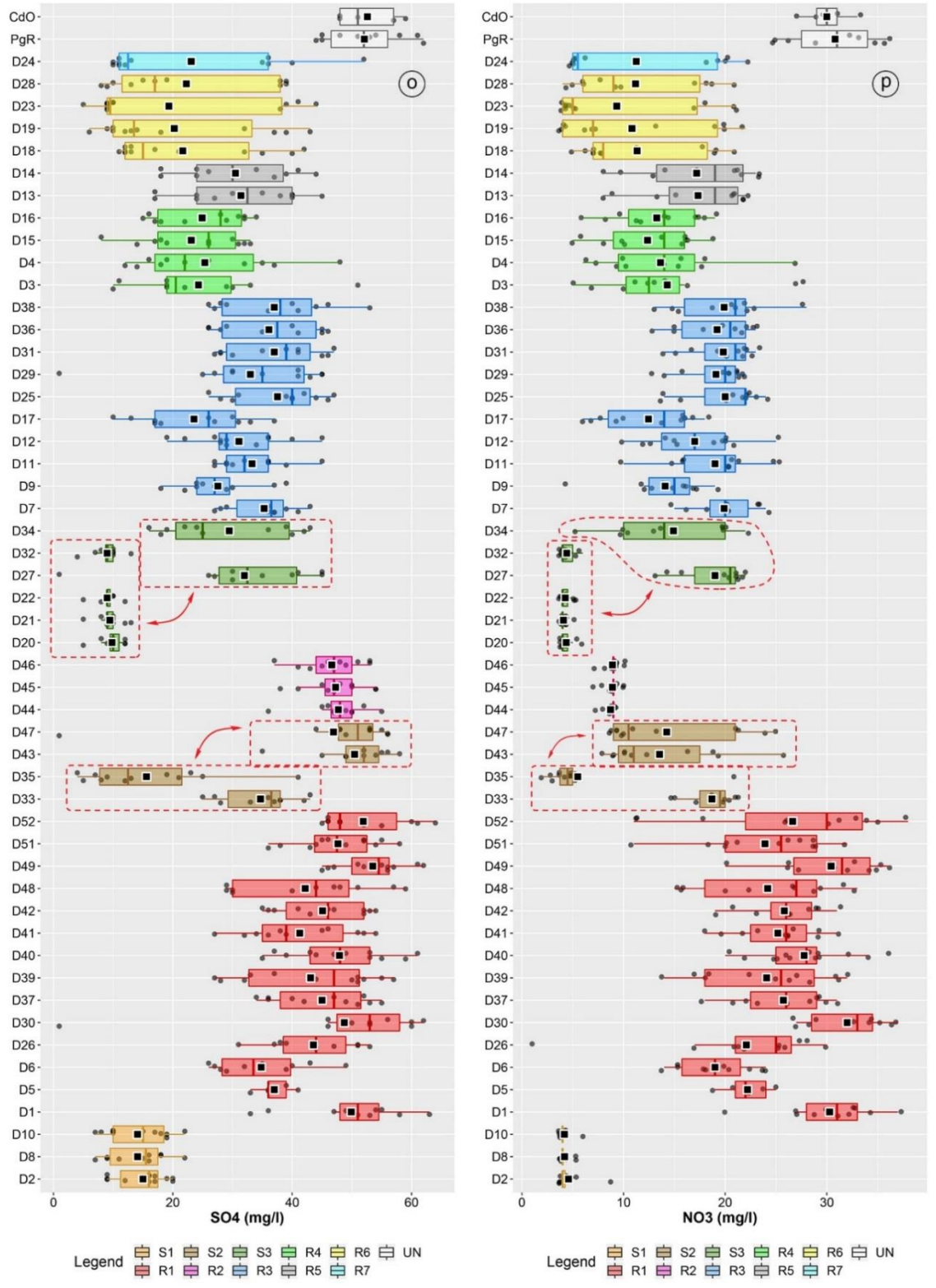


Fig. S4 (continued)

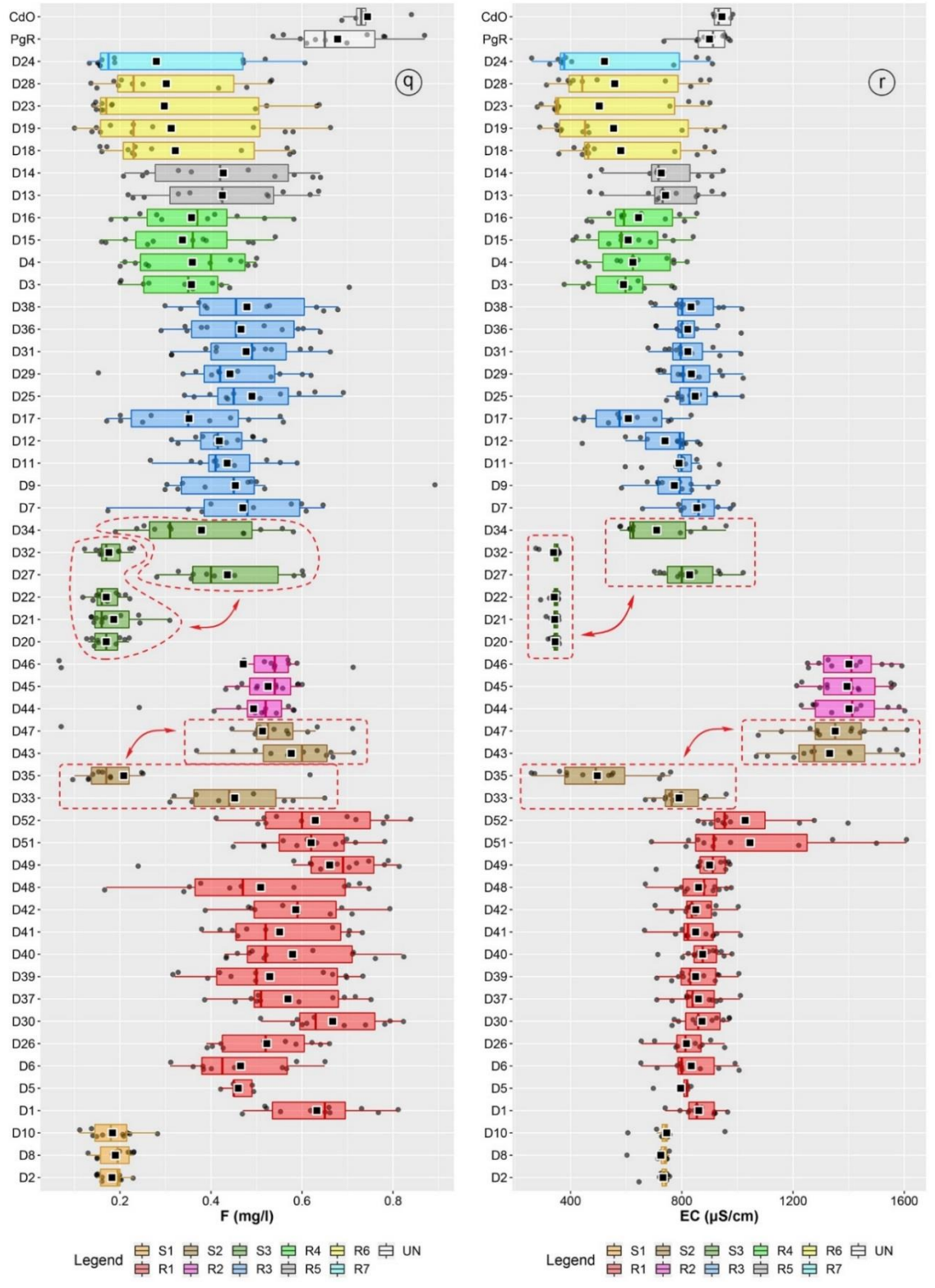


Fig. S4 (continued)

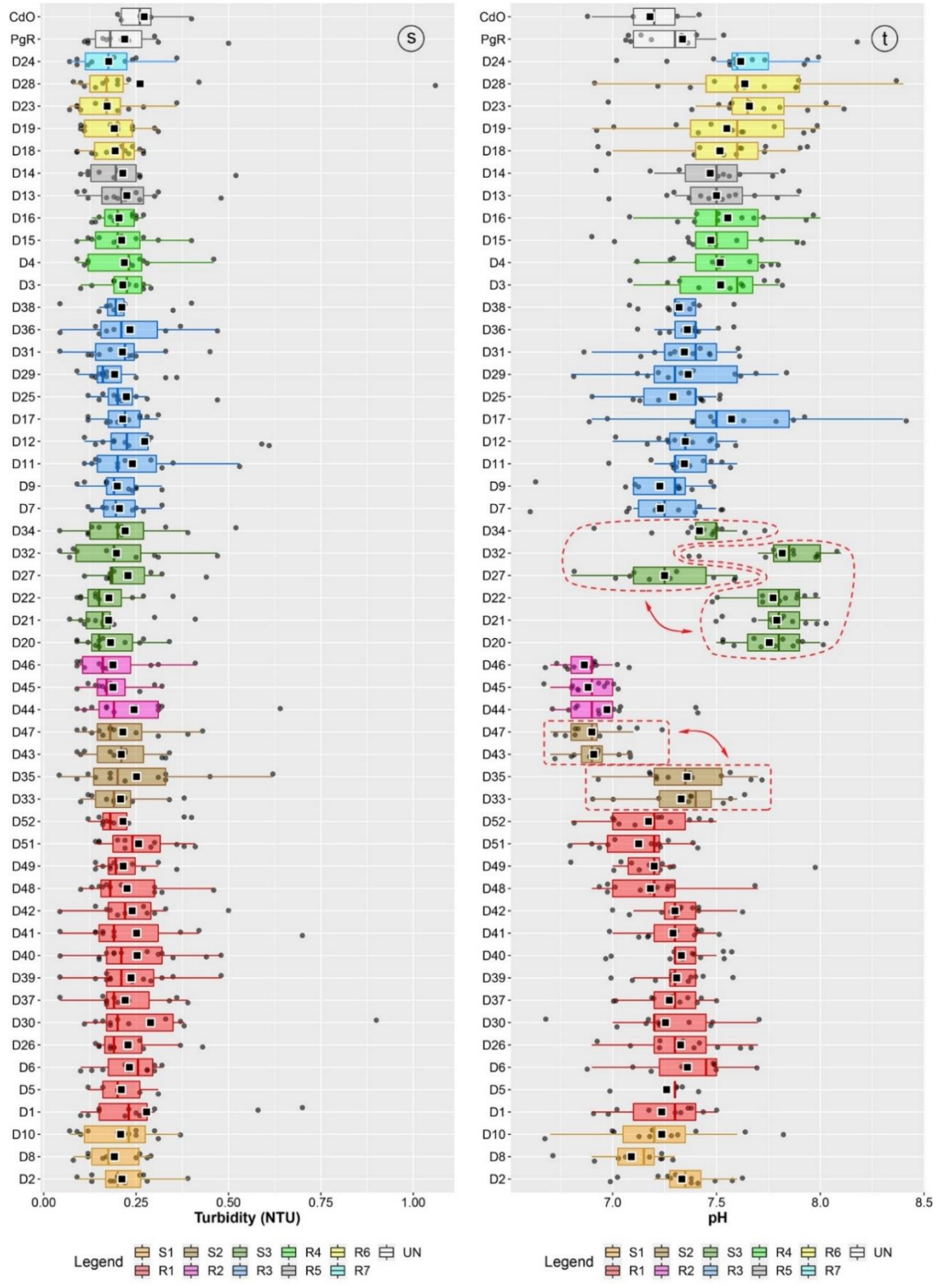


Fig. S4 (continued)

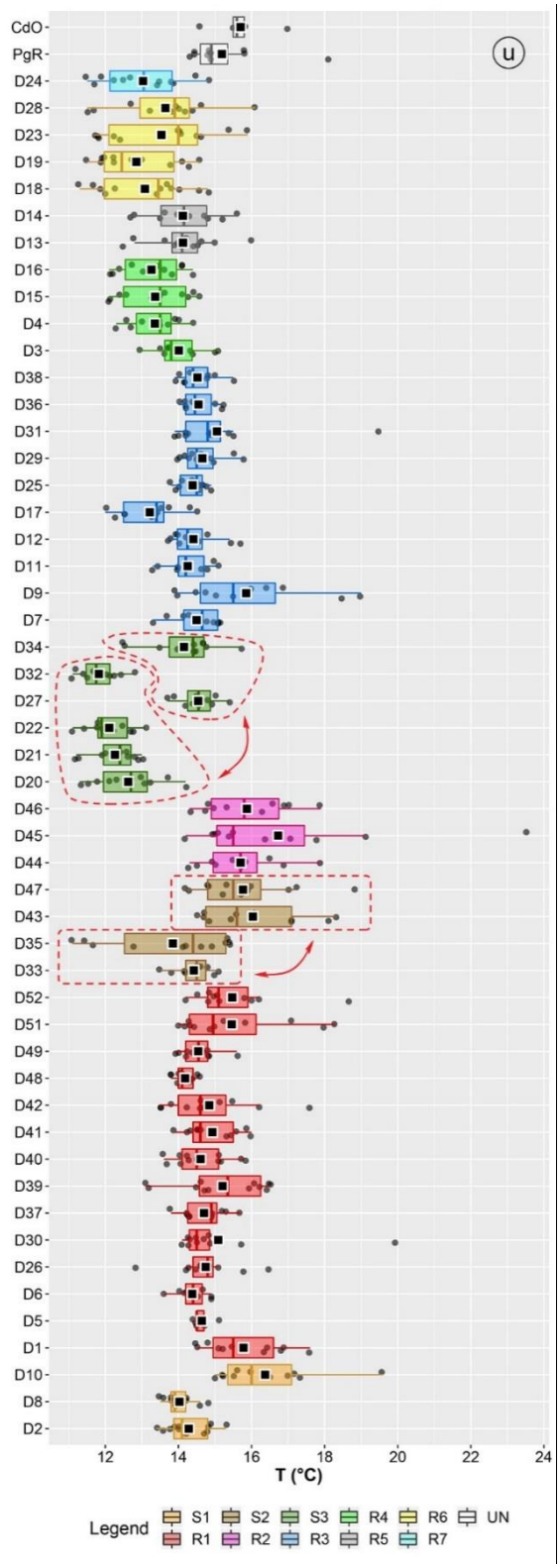


Fig. S4 (continued)

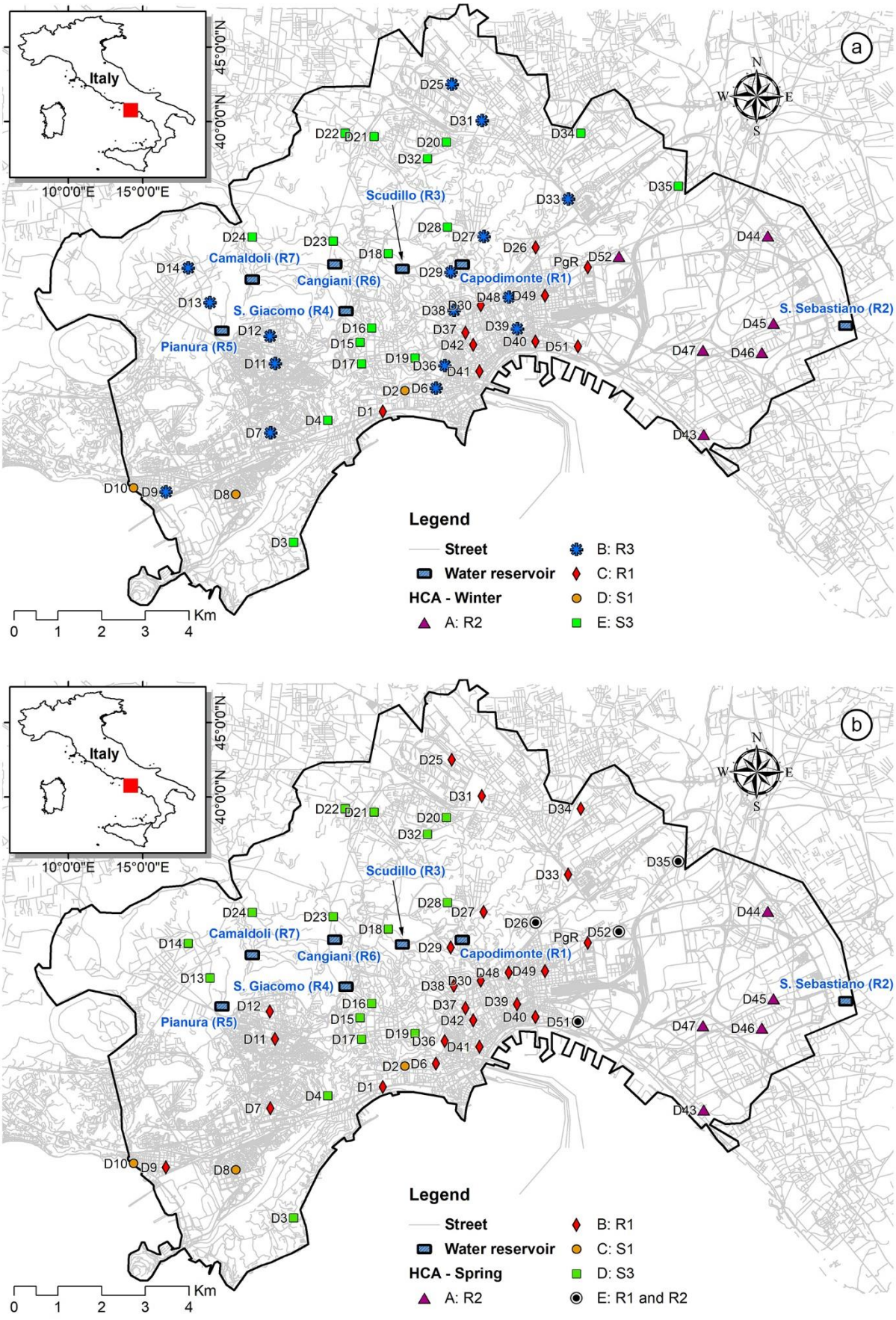
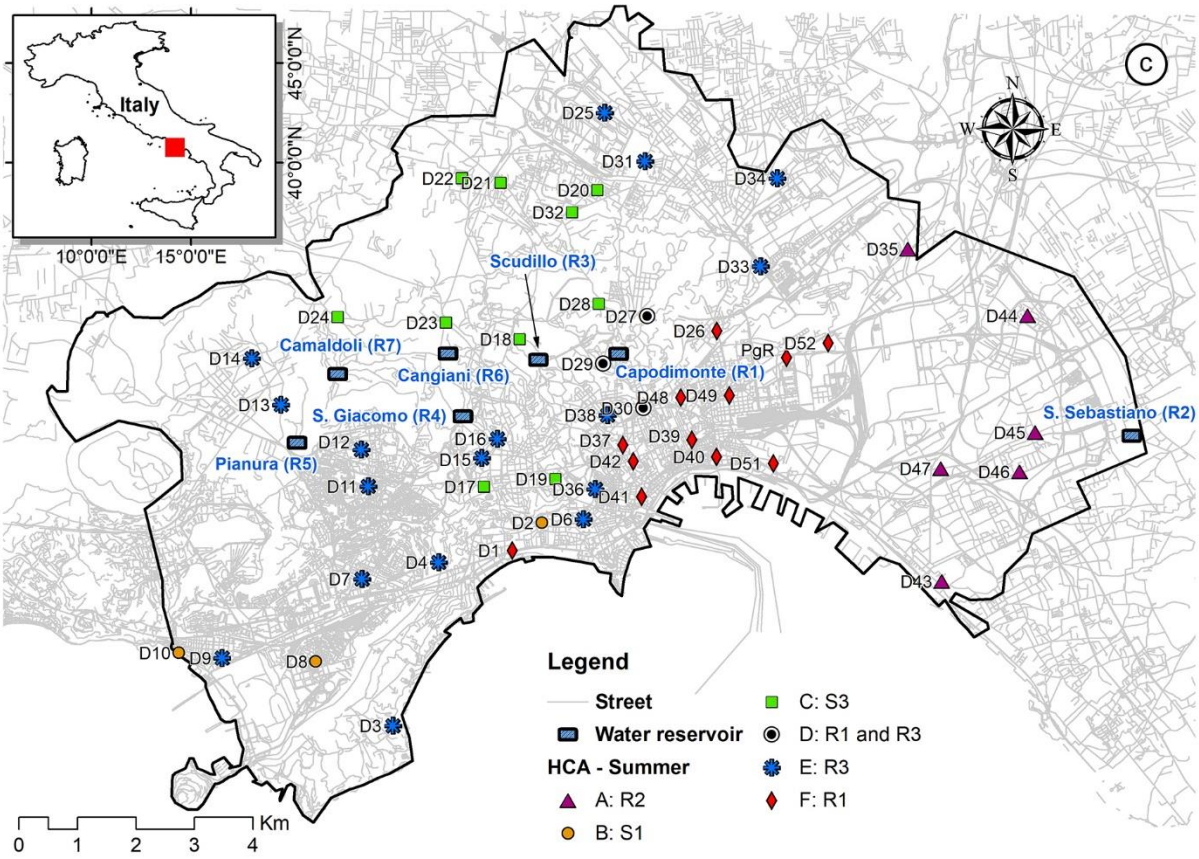


Fig. S5 Spatial representation of the results of seasonal hierarchical cluster analysis: (a) winter; (b) spring; (c) summer; and (d) fall. R1: Capodimonte; R2: S. Sebastiano; R3: Scudillo; S1: ACO; and S3: Serino Aqueduct.



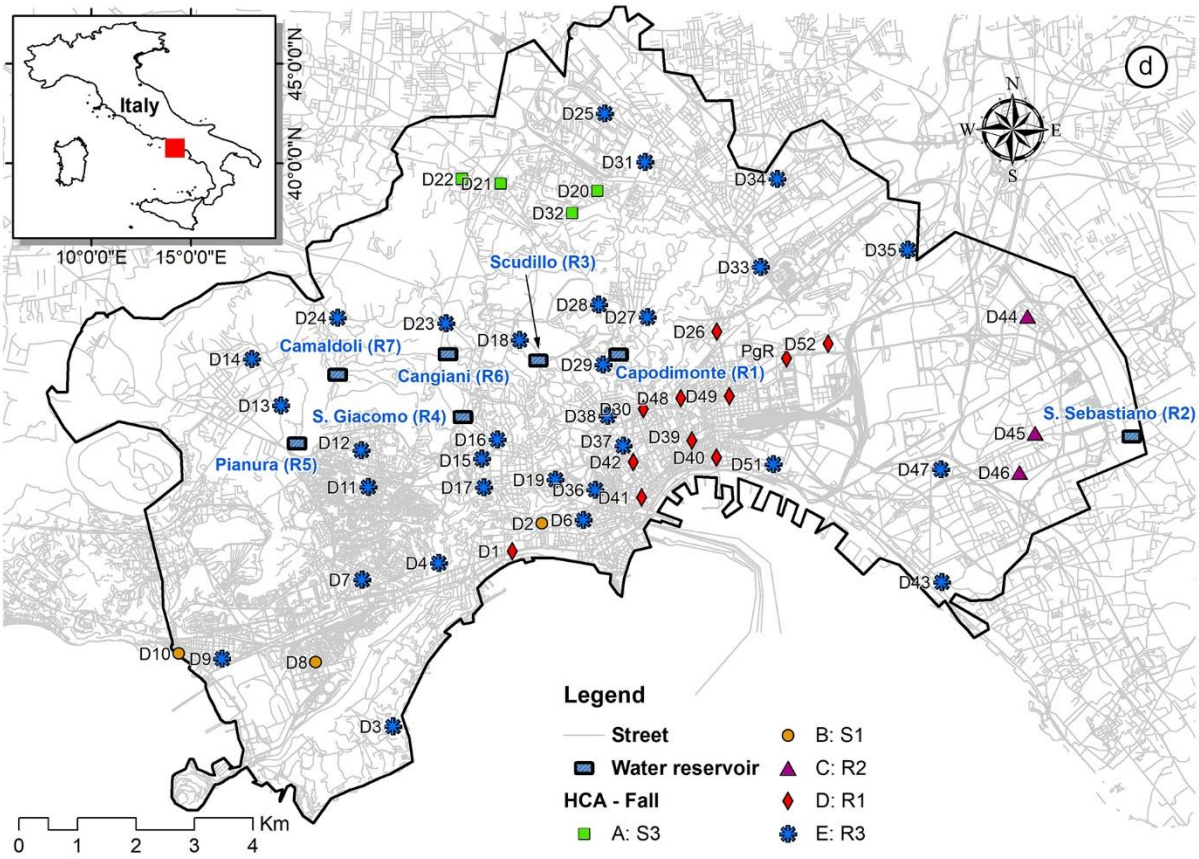


Fig. S5 (continued)

Text S1 HCA performance

According to the results and discussion in the manuscript, the following aspects are considered to evaluate performance of HCA combined with clr-transformation: 1) D6, D51 and D52 are excluded from R1, 2) D17 and D29 are excluded from R3, and D6 and D33 are included in R3, 3) D33 and D35 are excluded from S2, and 4) D27 and D34 are excluded from S3. Average Ca^{2+}, Mg^{2+}, Na^{+}, Cl^{-}, SO_4^{2-}, NO_3^{-} and F^{-} contents are then compared in the samples before and after (without and with asterisk, respectively, in Fig. S6) implementing the above-mentioned modifications. Figs. S6a and b show that the change in average chemical compositions of R1 and R3 is not clear. Hence, chemistry of the input waters is not represented for simplification. However, concentration of ions significantly increased in S2 after excluding D33 and D35. It is obvious that $S2^{*}$ is comparable with I3 and I4 indicating the highest contribution of these input waters and negligible importance of I5 in the period of this investigation (Fig. S6c). S3 is also plotted on I5 when D27 and D34 are ignored (Fig. S6d). Consequently, combination of HCA and clr-transformation works effectively with different sample sizes; however, capability of the approach can visually be verified when there are few samples.

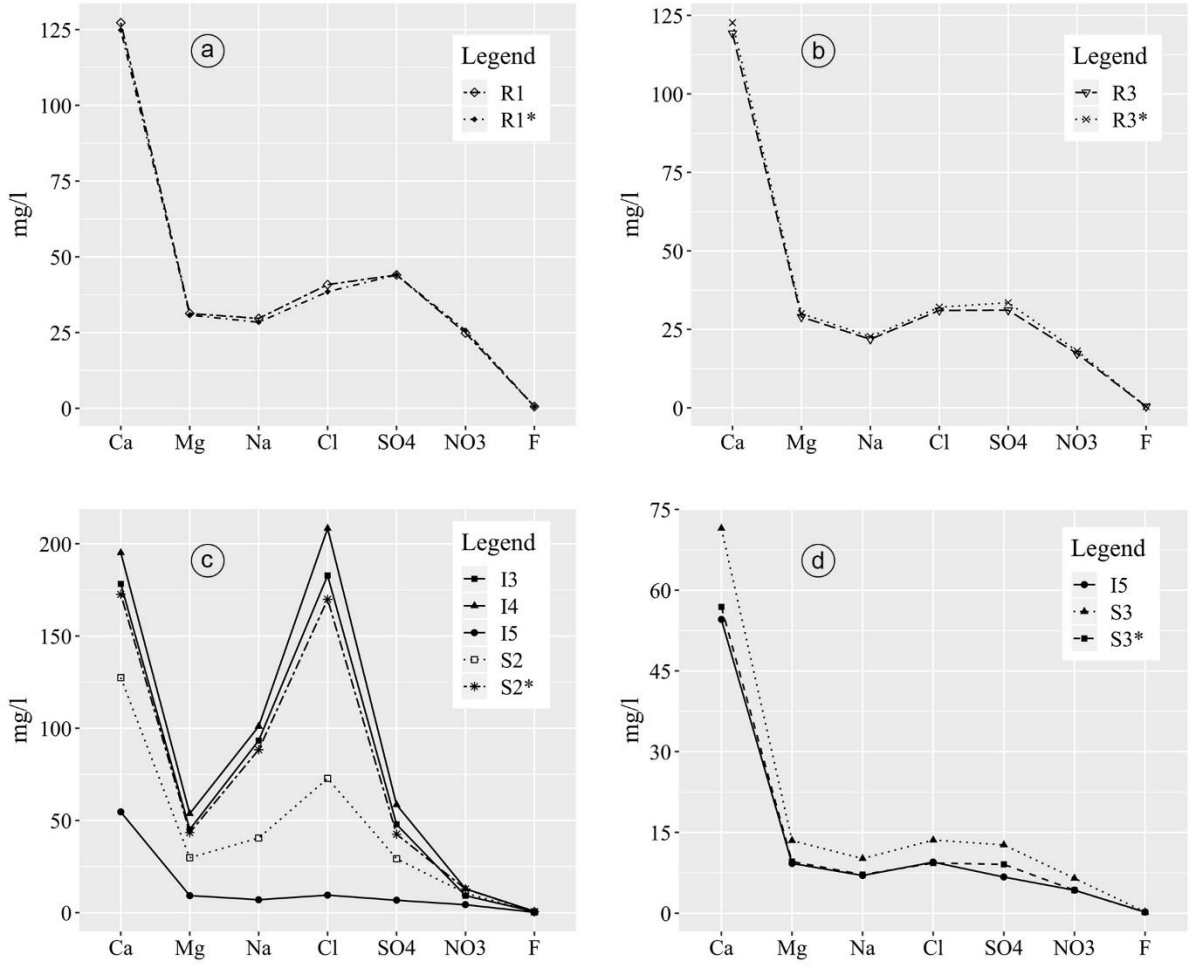


Fig. S6 The tap water composition before and after (without and with asterisk, respectively) implementing the HCA result: (a) R1: Capodimonte; (b) R3: Scudillo; (c) S2: Cancello well field, Campano Aqueduct and Ponte Tavano wells, I3: Cancello well field, I4: Campano Aqueduct and Ponte Tavano wells, and I5: Serino Aqueduct; and (d) S3: Serino Aqueduct.