

4. - ANALISI STATISTICA E DESCRITTIVA DEI DATI

4.1. - INTRODUZIONE

I dati analitici relativi alla composizione chimica delle acque dell'Arno analizzate sono stati trattati in maniera statistica. Le procedure statistiche, sia esplorative che inferenziali, sono qui proposte come metodologie volte a caratterizzare la natura stocastica dei sistemi naturali.

Il primo obiettivo di un'analisi statistica è quello della *sintesi delle informazioni* poiché di fronte all'analisi di fenomeni complessi, descritti da una pluralità di osservazioni, è necessario individuare parametri riassuntivi che in qualche modo rappresentino i dati nel loro complesso. Il secondo obiettivo è quello del riconoscimento di fattori, legami, leggi, non note in precedenza, e quindi da sottoporre a verifica mediante test di ipotesi. In generale, un'analisi statistica completa, su di una base di dati esaustiva, può essere riassunta da cinque fasi principali: 1) definizione degli obiettivi; 2) rilevazione dei dati; 3) elaborazione metodologica; 4) presentazione ed interpretazione dei risultati; 5) utilizzazione dei risultati della ricerca.

Nel nostro caso, gli obiettivi dell'indagine sono legati alla conoscenza delle caratteristiche dei dati riguardanti la geochimica delle acque di scorrimento superficiale del Bacino dell'Arno sia per quanto concerne il campionamento base che i monitoraggi successivi. La rilevazione dei dati è avvenuta applicando le metodologie classiche del campionamento delle acque, come riportato nel Capitolo 3. La matrice dei dati raccolti, (Appendice Tabelle A-H) mostra, mediante opportuni codici, le tipologie di acque campionate, ovverosia se appartenenti al corso principale dell'Arno o a suoi tributari, nonché le variabili fisiche e chimiche determinate. Le quantità che nella matrice sono precedute dal segno "<" sono associate, per quella variabile, in quello specifico campione, a tenori al di sotto del *limite di rilevabilità strumentale indicato*.

In una prima fase l'attenzione è stata rivolta a comprendere le caratteristiche chimico-fisiche relative al solo primo periodo di campionamento (maggio-agosto 2002 Tabella A in Appendice), campionamento che ha permesso la costituzione di una base di dati di partenza (campionamento 0). Successivamente, sono stati analizzati i dati raccolti durante i tre campionamenti (1, relativo al periodo gennaio-febbraio 2003, 2, relativo a maggio 2003 e 3, relativo a ottobre 2003; Tabelle B, C, D, in Appendice) dedicati al monitoraggio delle variazioni composizionali. I dati dei quattro perio-

di di campionamento (base più i tre periodi di monitoraggio) sono stati successivamente tra di loro confrontati al fine di evidenziare, per ogni variabile, le eventuali differenze e ipotizzarne la sorgente di perturbazione, mediante l'applicazione di test di ipotesi non parametrici.

La base di dati del campionamento di partenza (maggio-agosto 2002) è costituita da 203 osservazioni, di cui 40 rilevate nell'area del Casentino (13 nell'Arno, 27 nei tributari), 11 in Val di Chiana (1 nell'Arno, 10 nei tributari), 41 nel Valdarno Superiore (11 nell'Arno, 30 nei tributari), 11 nella Valle della Sieve (solo tributari), 38 nel Valdarno Medio (8 nell'Arno, 30 nei tributari) e 62 nel Valdarno Inferiore (17 nell'Arno, 45 nei tributari). Complessivamente, 50 campioni sono relativi, nei vari bacini, al Fiume Arno, mentre 153 ai suoi tributari principali. Nel corso dei tre periodi di monitoraggio sono stati raccolti 150 campioni (50 per ogni campionamento), di cui 30 sono relativi al Casentino, 8 alla Val di Chiana, 18 al Valdarno Superiore, 6 alla Val di Sieve, 28 al Valdarno Medio e 60 al Valdarno Inferiore. Complessivamente, 70 campioni sono stati prelevati lungo il corso del Fiume Arno mentre 80 sono relativi ai suoi tributari.

Infine, è stata effettuata una indagine volta alla conoscenza preliminare della distribuzione di numerose specie metalliche e non, quali: Al, As, Ba, Cd, Cr, Cu, Fe, Li, Mn, Mo, Ni, P, Pb, Rb, Se, U, V, W, Y, Zn, per un totale di 22 misure (Tabelle E, F, G e H in Appendice), di cui 10 relative all'asta dell'Arno e 12 ai suoi tributari (3 in Casentino, 2 in Val di Chiana, 3 nel Valdarno Superiore, 1 nella Sieve, 2 nel Valdarno Medio e 11 nel Valdarno Inferiore).

4.2. - ANALISI DELLA BASE DI DATI ORIGINARIA

I parametri analizzati dal punto di vista statistico per la base di dati di partenza sono pH (200 casi) e TDS (203 casi), oltre le variabili chimiche (specie maggiori, minori e specie azotate): $\text{HCO}_3^- + \text{CO}_3^{2-}$, Cl^- , SO_4^{2-} , Na^+ , K^+ , Ca^{2+} , Mg^{2+} , F^- , Br^- , B , NH_4^+ , NO_2^- , NO_3^- (203 misure), SiO_2 (202 misure). Per ognuna di queste componenti sono state costruite delle Tavole (Tavole 1a-16a) nelle quali sono riportati: 1) un diagramma quantile-quantile per ogni variabile, come strumento per analizzare la forma della distribuzione di frequenza, individuare presenza di gruppi di dati distinti e/o valori anomali; 2) una *dot map* dei valori tramite la quale si associa al tenore crescente della variabile il luogo di campionamento all'interno del bacino; 3) un *box-plot*, al fine di comparare per ogni variabile i valori rilevati nei vari bacini.

Quando sono risultati presenti, per una variabile, valori inferiori al limite di rilevabilità, questi sono stati sostituiti, come è prassi usuale, con il valore del limite stesso, quando in numero basso rispetto al totale delle osservazioni. Nelle tavole, oltre ad una sintesi del comportamento della variabile, o specie chimica analizzata sono stati anche riportati i valori massimo, minimo e dei percentili del 25, 50 (mediana) e 75.

Con l'intento di facilitare l'interpretazione dei risultati ricordiamo che in statistica, per un insieme di dati ordinati in modo non decrescente, il quantile è un valore che lascia al di sopra, o al di sotto, una determinata percentuale di osservazioni. In tali diagrammi i quantili sperimentali ottenuti dai dati (cerchi pieni rossi) sono confrontati con i quantili di variabili aleatorie prese come modello di riferimento (la linea continua nera), quali ad esempio la Normale di Gauss o la Log-Normale. Ricordiamo inoltre che la *mediana*, indicata con Me , è definita come quel valore di una variabile X che bipartisce la distribuzione ordinata delle modalità $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Quindi, la *mediana* è determinata in modo che metà delle osservazioni siano ad essa inferiori e metà superiori; di conseguenza la *mediana* è la modalità della unità statistica che occupa il posto centrale nella distribuzione ordinata delle osservazioni. La *mediana* gode di alcune proprietà, tra le quali, quella sicuramente più importante, è la sua capacità di essere rappresentativa della posizione della distribuzione anche in presenza di valori estremi notevolmente diversi da tutti gli altri; tale requisito è detto *resistenza* e costituisce un indubbio vantaggio di questo indice rispetto alla *media*, soprattutto in distribuzioni di frequenza così asimmetriche come quelle riscontrate nella base di dati analizzata. Infine, i diagrammi a scatola (*box-plot*), introdotti da TUKEY nel 1977, sono basati sulla determinazione dei *quantili* e dei valori *massimo* e *minimo* di una variabile. Tali indicatori sono rappresentati, rispetto ad un conveniente asse, mediante una scatola che si estende tra il *primo* ed il *terzo quantile*; all'interno di tale scatola è messa in evidenza la posizione della mediana della distribuzione. Alle due estremità della scatola si pongono delle linee - note anche come *baffi* o *whiskers* - che raggiungono i due estremi (*massimo* e *minimo*) della distribuzione, se inclusi in un intervallo commisurato alla differenza *interquartile* ($Q_3 - Q_1$); in caso contrario gli estremi (valori *anomali* o *outliers*) si evidenziano con appositi simboli (nel nostro caso dei +). L'utilità del *box-plot* deriva dal fatto che, in una sola ed essenziale rappresentazione grafica, si evidenziano con indici *robusti*, tutti gli aspetti di una distribuzione di frequenza ritenuti importanti. Il

box-plot, indica infatti la posizione della variabile mediante la *mediana* e la variabilità mediante l'ampiezza della scatola (la *differenza interquartile*); inoltre, il posizionamento della mediana all'interno della scatola, anche rispetto agli estremi della distribuzione, mostra l'eventuale *asimmetria* e la sua direzione. Infine, trattandosi di una rappresentazione grafica semplice, il *box-plot* permette un immediato confronto tra più variabili o tra la stessa variabile esaminata in tempi, luoghi e situazioni differenti.

Nella Tavola I (a) è riportata l'analisi statistica delle misure di pH. Il valore più basso registrato nel bacino è pari a 6.62, quello più alto a 9.73. Come si nota dal diagramma dei quantili (in alto a destra) la distribuzione di frequenza dei dati segue approssimativamente la curva normale, fatta eccezione per alcuni valori anomali che si trovano nelle code, per valori < 7.0 e valori > 9.0 . La *dot map* mostra che i valori più bassi sono stati in prevalenza rilevati nella parte occidentale del bacino; in questo contesto i valori elevati o bassi appaiono essere associati a situazioni puntuali. I diagrammi a scatola mostrano le variazioni di pH nei vari bacini (CA = Casentino, CH = Val di Chiana, VAS = Valdarno Superiore, SI = Bacino della Sieve, VAM = Valdarno medio, VAI = Valdarno Inferiore) e permettono di osservare che i valori mediani maggiori sono associati al Casentino mentre quelli più bassi al Valdarno inferiore.

Nella Tavola II (a) è riportata l'analisi statistica della salinità (o TDS). Il valore minore registrato è pari a 83 mg/L, quello maggiore a 13272 mg/L. Nel diagramma dei quantili, la distribuzione di frequenza dei dati segue approssimativamente la curva normale, una volta adottata la trasformazione logaritmica, fatta eccezione per i dati i cui valori sono maggiori di circa 1800 mg/L. Il fatto che i dati seguano una legge log-normale rivela la presenza di un meccanismo di proporzionalità che lega i valori, privilegiando in frequenza quelli bassi, così come accade ad esempio per la conducibilità. La *dot map* mostra, com'era prevedibile, la tendenza ad un aumento dei valori di salinità dalle aree sorgive a quelle di pianura, con un ulteriore deciso incremento per la zona vicina alla foce, dove si risente dell'intrusione marina, nonostante che i campioni di Arno Vecchio (VAI XLIX) e Bocca d'Arno (VAI L) non siano stati inseriti perché troppo elevati. Si rileva ancora l'incremento dei valori della TDS nella porzione meridionale del bacino. I diagrammi a scatola mostrano come la variabilità dei valori della TDS sia maggiore nei bacini del Valdarno Inferiore, Medio e della Chiana.

La Tavola III (a) riporta l'analisi statistica dei valori di $\text{HCO}_3^- + \text{CO}_3^{2-}$. Il valore minore registrato è pari a 42 mg/L, quello maggiore a 836 mg/L (diagramma dei quantili). La distribuzione di frequenza dei dati segue approssimativamente la curva normale, fatta eccezione per un gruppo di dati i cui valori sono minori di circa 150 e maggiori di 400 mg/L. Il fatto che anche in questo dato i dati tendano a seguire la legge log-normale rivela che i meccanismi principali all'origine dei valori registrati sono regolati da fenomeni di diluizione/concentrazione. La *dot map* mostra la presenza di una zona settentrionale al fiume Arno piuttosto omogenea e caratterizzata dai valori più bassi. Dai diagrammi a scatola si rileva che il valore mediano più elevato in HCO_3^- è relativo alla Val di Chiana, quello più basso al Casentino. Da notare la scarsa variabilità dei dati nel Bacino della Sieve.

Nella Tavola IV (a) è riportata l'analisi dell'ione Cl^- . Il valore più basso è pari a 4.6 mg/L, il più alto a 7500 mg/L. Dalla segmentazione presente nel diagramma dei quantili, anche dopo aver adottato la trasformazione logaritmica, si notano più gruppi di dati dal comportamento distinto, localizzati nelle code della distribuzione. La *dot map* mostra una netta distinzione tra la zona settentrionale e orientale del bacino nei confronti di quella meridionale e l'incremento marcato nei pressi della foce. Dai diagrammi a scatola si rileva come il valore mediano più elevato sia chiaramente associabile con il Valdarno Inferiore così come la maggior presenza di dati anomali.

Nella Tavola V (a) è riportata l'analisi statistica dei valori di SO_4^{2-} . Il valore minore registrato è pari a 6.70 mg/L, quello maggiore a 1275 mg/L. La disposizione dei punti nel diagramma dei quantili è simile a quella del Cl^- , con valori anomali nelle code della distribuzione (rispettivamente <20 mg/L e >150 mg/L). La *dot map* mostra, anche in questo caso, la diversità di distribuzione tra la parte settentrionale e quella meridionale così come l'incremento dei tenori nei pressi della foce. Dai diagrammi a scatola si rileva come il valore mediano più elevato sia chiaramente associabile con il sottobacino del Valdarno Inferiore così come la maggior presenza di dati anomali.

Nella Tavola VI (a) è riportata l'analisi statistica dei valori di Na^+ . Il valore minore registrato è pari a 4.10 mg/L, quello maggiore a 3980 mg/L. La disposizione dei punti nel diagramma dei quantili è simile a quella di Cl^- e di SO_4^{2-} , con ancora valori anomali nelle code (< 7.5 mg/L e > 150 mg/L). La *dot map* evidenzia, come nei casi precedenti, la differenza tra l'area settentrionale, dove sono presenti i tenori più bassi, e quella

meridionale, caratterizzata da una maggiore variabilità. Anche il Na^+ , così come Cl^- e SO_4^{2-} , cresce in modo marcato nei pressi della foce. Dai diagrammi a scatola si rileva come il valore mediano più elevato di Na^+ sia quello del Valdarno Inferiore, seguito dalla Val di Chiana.

Nella Tavola VII (a) è riportata l'analisi statistica del K^+ . Il valore minore è pari a 0.7 mg/L, quello maggiore a 153 mg/L. Il diagramma dei quantili è compatibile con la presenza di una distribuzione log-normale dei valori, con casi anomali nelle code (< 1 mg/L e > 20 mg/L). La *dot map* mostra ancora una chiara differenza tra la regione settentrionale ed orientale, dove si registrano i valori più bassi, ed una maggiore variabilità in quella centro-meridionale. Nei *box plot* si rileva come il valore mediano più elevato sia relativo alla Val di Chiana e al Valdarno Inferiore, dove è presente anche il maggior numero di dati anomali.

Nella Tavola VIII (a) è riportata l'analisi statistica del Ca^{2+} . Il valore minore registrato è pari a 14 mg/L, quello maggiore a 471 mg/L. La disposizione dei punti nel diagramma dei quantili è compatibile con la presenza di una distribuzione log-normale, con presenza di valori anomali nelle code (< 20 mg/L e > 150 mg/L). La *dot map* che associa il valore di Ca^{2+} alle coordinate del punto di prelievo mostra, rispetto alle variabili precedenti, una sostanziale omogeneità nell'intervallo 10-200 mg/L, fatta eccezione per alcuni punti lungo il medio e basso corso dell'Arno e nell'area meridionale. Dai diagrammi a scatola si rileva che il valore mediano più elevato sia relativo alla Val di Chiana, mentre la variabilità maggiore compete al Valdarno Inferiore. Scarsa risulta invece la variabilità in Val di Sieve.

La Tavola IX (a) riporta l'analisi statistica del Mg^{2+} dalla quale si rileva che 1.5 mg/L è il valore più basso e 490 mg/L quello più elevato. La disposizione dei punti nel diagramma dei quantili è compatibile con la presenza di una distribuzione log-normale dei valori con presenza di casi anomali nella coda destra (valori > 55 mg/L). La *dot map* mostra la presenza di una certa variabilità che caratterizza non solo il corso del Fiume Arno, ma anche la zona sud-occidentale rispetto a quella settentrionale. Da notare, anche in questo caso, la presenza di valori elevati nei pressi della foce, un incremento che sarebbe ancora più marcato se i campioni di Arno Vecchio (VAI XLIX) e Bocca d'Arno (VAI I) fossero stati inseriti nell'analisi spaziale. Dai diagrammi a scatola si osserva come il valore mediano più elevato di Mg^{2+} sia associato alla Val di Chiana mentre la variabilità maggiore, così come la presenza di valori anomali, compete anche in questo caso al Valdarno Inferiore.

Nuovamente scarsa è la variabilità dei dati raccolti lungo la Sieve.

Nella Tavola X (a) è riportata l'analisi statistica dei valori di NH_4^+ . Il valore minore registrato è pari a 0.01 mg/L, quello maggiore a 10.45 mg/L. La disposizione dei punti nel diagramma dei quantili è compatibile con almeno due gruppi di dati, caratterizzati da valori minori e maggiori di circa 0.4 mg/L. La *dot map* mostra la presenza di una certa omogeneità, talvolta interrotta da valori puntuali elevati, localizzati spesso lungo il corso dell'Arno o della Chiana. I diagrammi a scatola rilevano come il valore mediano più elevato di NH_4^+ sia decisamente nel bacino della Chiana.

La Tavola XI (a) mostra l'analisi dei valori di NO_2^- . Il valore minimo registrato è 0.006 mg/L, quello massimo 11.5 mg/L. La disposizione dei punti nel diagramma dei quantili è compatibile con la presenza di disomogeneità nei dati per valori < 0.02 mg/L. La *dot map* mostra la tendenza dei valori a crescere dalle zone orientali verso quelle occidentali e da quelle settentrionali verso le meridionali. I diagrammi a scatola rilevano che i valori mediani più elevati si trovano nel Valdarno Inferiore, il Valdarno Medio e nella Chiana; da notare i valori elevati del Valdarno Inferiore e Medio ed i valori particolarmente bassi del Casentino e della Val di Sieve.

Nella Tavola XII (a) è riportata l'analisi statistica dei valori di NO_3^- . Il valore minore registrato è pari a 0.01 mg/L, quello maggiore a 72.5 mg/L; quattro valori risultano inferiori al limite di rilevanza (< 0.01 mg/L). La disposizione dei punti nel diagramma dei quantili è compatibile con la presenza di una distribuzione log-normale con casi anomali per valori < 0.14 mg/L e > 12 mg/L. La *dot map* mostra la presenza di un'ampia variabilità con un incremento dei valori dalle aree sorgive verso quelle di pianura. I diagrammi a scatola rilevano come il valore mediano più elevato di NO_3^- sia associato con il Valdarno Inferiore mentre la variabilità maggiore spetta alla Val di Chiana. Da notare la presenza di valori particolarmente bassi nel Casentino.

Nella Tavola XIII (a) è riportata l'analisi statistica dei valori di F^- . Il valore minore registrato è pari a 0.03 mg/L, quello maggiore a 2.0 mg/L. La disposizione dei punti nel diagramma dei quantili è compatibile con la presenza di una distribuzione log-normale con casi anomali per valori < 0.08 mg/L. La *dot map* mostra la presenza di un certo incremento nei valori e nella variabilità dalle aree nord-orientali verso quelle sud-occidentali. I valori minori sono, in ogni caso, registrati in Casentino. I diagrammi a scatola rilevano il valore mediano più elevato in Valdarno Inferiore così come la variabi-

lità maggiore, seguito dalla Chiana e dal Valdarno Medio. Da notare anche in questo caso la presenza di valori estremamente bassi nel Casentino.

Nella Tavola XIV (a) è riportata l'analisi statistica dei valori di Br^- . Il valore minore registrato è pari a 0.001 mg/L, quello maggiore a 27.5 mg/L. La disposizione dei punti nel diagramma dei quantili è compatibile con la presenza di una distribuzione log-normale con casi anomali per valori > 7.4 mg/L. La *dot map* mostra la tendenza, già registrata per altre variabili, alla crescita dei valori dalle aree sorgive a quelle di pianura, lungo il corso dell'Arno e nell'area meridionale rispetto a quella settentrionale. L'area della foce, influenzata dalla intrusione marina, mostra i valori più elevati, nonostante l'assenza dei campioni di Arno Vecchio (VAI XLIX) e Bocca d'Arno (VAI L). I diagrammi a scatola rilevano come il valore mediano più elevato di Br^- sia associato con il Valdarno Inferiore così come la variabilità maggiore e la presenza di valori anomali elevati.

Nella Tavola XV (a) è riportata l'analisi statistica dei valori di SiO_2 . Il valore minore è pari a 0.30 mg/L, il maggiore a 34.6 mg/L. La disposizione dei punti nel diagramma dei quantili è log-normale con casi anomali per valori < 4.5 mg/L e > 20 mg/L. La *dot map* mostra la presenza di variabilità che appare caratterizzare il medio e basso corso dell'Arno così come l'area sud-occidentale. I diagrammi a scatola rilevano come il valore mediano più elevato di SiO_2 sia associato con il Valdarno Inferiore e come la variabilità sia comunque elevata in tutta l'area del Valdarno.

Nella Tavola XVI (a) è riportata l'analisi statistica del B. Il valore minore registrato è pari a 0.003 mg/L, quello maggiore a 6.3 mg/L. La disposizione dei punti nel diagramma dei quantili è compatibile con la presenza di più gruppi di dati: < 0.02 mg/L, compresi tra 0.02 e 1 mg/L e > 1 mg/L. La *dot map* mostra la tendenza dei valori a crescere dall'area settentrionale e orientale verso il medio e basso corso dell'Arno, così come verso l'area meridionale. I diagrammi a scatola rilevano come il valore mediano più elevato di B sia associato con il Valdarno Inferiore e Medio, mentre bassi sono i valori registrati nel Casentino e nella Sieve.

4.3. - LE VARIAZIONI CHIMICHE TEMPORALI E SPAZIALI

Al fine di evidenziare per ogni variabile le eventuali differenze ed ipotizzarne la sorgente di perturbazione, i dati raccolti durante i tre campionamenti dedicati al monitoraggio delle variazioni

composizionali sono stati confrontati con quelli del campionamento base. In tabella 8 è riportata una cross-tabulazione tra la data del campionamento e la sigla con la quale sono stati etichettati i vari periodi di prelievo per $i = 0, \dots, 3$, dove con il numero 0 è indicato il campionamento relativo alla base di dati di partenza, con 1 il periodo di monitoraggio gennaio-febbraio 2003, 2 quello

relativo a maggio 2003 e, infine, con 3 il periodo conclusivo di ottobre 2003.

Tale tabella permette di visualizzare in modo chiaro e semplice quanti campioni siano stati raccolti per ogni data di prelievo. Le 351 analisi, divise in quattro gruppi (201 per il campionamento base, 50 per ognuno dei campionamenti 1, 2 e 3), sono state confrontate utilizzando, come metodo

Tab. 8 - *Cross-tabulazione tra la data di campionamento e i vari periodi di prelievo dei campioni nel Bacino dell'Arno; con il numero zero è indicato il campionamento volto a costituire la base di dati di partenza.*

- Cross-tabulation between the May-August 2002 sampling dates and those of the monitoring (January-February, May and October 2003) periods; "0" refers to the May-August 2002 sampling, whereas "1", "2" and "3" are for the January-February, May and October 2003, respectively.

Data campionamento	Monitoraggio 0 (base)	Monitoraggio 1	Monitoraggio 2	Monitoraggio 3	Totale campioni prelevati
18-06-2002	13				13
18-02-2002	7				7
02-05-2002	9				9
09-05-2002	19				19
23-05-2002	18				18
05-06-2002	13				13
24-06-2002	18				18
23-07-2002	16				16
30-07-2002	18				18
19-08-2002	14				14
20-08-2002	10				10
01-10-2002	20				20
08-10-2002	17				17
23-01-2003	1	11			12
27-01-2003	5	9			14
12-02-2003	2	13			15
20-02-2003	1	14			15
26-02-2003		3			3
20-05-2003			18		18
26-05-2003			19		19
27-05-2003			13		13
24-10-2003				19	19
25-10-2003				9	9
26-10-2003				19	19
02-11-2003				1	1
data non segnata				2	2
	201	50	50	50	351

grafico, *box-plot* comparativi (Tavole 1b-16b), e come metodo statistico il test non parametrico della mediana.

Il test della mediana effettua un confronto tra il valore di questa statistica in ogni gruppo con quello ottenuto considerando i dati come tutti appartenenti ad una unica popolazione. Se la significatività del test risulta essere inferiore a 0.05, l'ipotesi nulla di una mediana comune a tutti i gruppi deve essere scartata; vi saranno pertanto, per almeno un gruppo, differenze significative. La scelta di utilizzare un test non parametrico come quello della mediana si basa su alcune considerazioni relative alla presenza di distribuzioni di frequenza fortemente asimmetriche così come di numerosi valori anomali, come si è già avuto

Tab. 9 - Risultati del test della mediana (χ^2 , sono i valori della distribuzione campionaria chi-quadrato; gl, sono i gradi di libertà; Sig, è il valore della significatività statistica); in blu le variabili per le quali non si sono evidenziate differenze significative, in nero quelle con significatività tra 0.01 e 0.05 ed infine, in rosso, quelle con significatività inferiore o uguale a 0.01.

- Results of the median test (χ^2 are the values of the chi-square sample distribution; gl, are the degree of freedom; Sig, are the significance values); the variables without significant differences are reported in blue; variables with significance between 0.01 and 0.05 are in black, whereas in red are reported the variables with significance equal or < 0.01.

	Numerosità	Mediana	(χ^2)Chi-quadrato	gl	Sig.
Temperatura	347	17.50	134.258	3	0.000
pH	337	8.05	14.203	3	0.003
Cond.(mS/cm)	349	0.60	3.564	3	0.313
HCO ₃ +CO ₃	349	247.00	5.842	3	0.120
Cl	349	32.260	8.804	3	0.032
SO ₄	349	37.500	7.926	3	0.048
NH ₄	349	0.150	10.206	3	0.017
NO ₂	346	0.090	11.280	3	0.010
NO ₃	349	4.250	15.600	3	0.001
Na	349	32.600	11.693	3	0.009
K	349	2.940	11.974	3	0.007
Ca	349	69.00	1.072	3	0.784
Mg	349	13.500	4.797	3	0.187
F	349	0.230	11.357	3	0.010
Br	349	0.100	14.418	3	0.002
SiO ₂	346	6.250	5.132	3	0.162
B	349	0.090	21.724	3	0.000
TDS	349	459.300	3.752	3	0.290

modo di discutere in precedenza. In tali situazioni un test del genere è sicuramente più adeguato per l'obiettivo che si vuole raggiungere, e cioè evidenziare, se presenti, differenze significative tra il periodo base e quelli successivi.

Dai risultati del test, riportati in tabella 9 emerge che variabili quali salinità (TDS), la somma (HCO₃⁻+CO₃²⁻), Ca²⁺, Mg²⁺ e SiO₂, non mostrano differenze significative tra il campionamento 0 e i periodi di monitoraggio 1, 2 e 3. Evidentemente tali variabili risultano essere influenzate da vari fattori capaci, nel corso del tempo, di far variare il valore nell'intorno di una mediana comune. Variabili come Cl⁻, SO₄²⁻ e NH₄⁺ mostrano invece differenze significative tra i vari periodi di monitoraggio, differenze che possono essere analizzate in dettaglio mediante lo studio dei diagrammi a scatola riportati nelle Tavole IV (b), V (b) e X (b) e dei risultati numerici della tabella 10, dove sono riportati i valori minimo, massimo e mediano, sia separatamente per periodo di campionamento sia considerando tutti i periodi di prelievo come appartenenti ad una unica popolazione.

Dai grafici si rileva come per lo ione Cl⁻ sia il monitoraggio 1 a differenziarsi rispetto agli altri a causa dei valori di più bassa concentrazione; quest'ultimi variano infatti entro un intervallo compreso tra 3 e 1874 mg/L, mentre negli altri periodi i valori sono compresi tra 4 e 7500 mg/L. L'analisi dei risultati della variabile SO₄²⁻ (Tavola 5b) rileva che anche in questo caso è il campionamento 1 che tende a presentare valori inferiori rispetto alla mediana generale (39.84 mg/L). Al contrario, nel caso della variabile NH₄⁺, tutti i periodi di campionamento mostrano differenze sia nel valore mediano che nella presenza di dati anomali. Rispetto a Cl⁻ e SO₄²⁻, quindi, NH₄⁺ si rileva una variabile molto sensibile al momento del campionamento.

Il risultato del test della mediana (*Me*) su variabili quali pH, NO₂⁻, NO₃⁻, Na⁺, K⁺, F⁻ e Br⁻ rivela la presenza di differenze significative per valori di $\alpha < 0.01$. I grafici dei diagrammi a scatola sono riportati, rispettivamente, nelle Tavole I (b), X (b), XII (b), VI (b), VII (b), XIII (b) e XVI (b), mentre i valori minimo, massimo e mediano sono indicati in tabella 11 sia separatamente, per periodo di campionamento, che considerando tutti i periodi di prelievo come appartenenti ad un'unica popolazione. L'effetto discriminante del campionamento 1 si riscontra per le variabili pH (*Me* = 7.93) ed NO₂⁻ (*Me* = 0.046 mg/L), per le quali sono registrati i valori inferiori. Nel caso di NO₃⁻, invece, sono i campionamenti 0 (*Me* = 3.41 mg/L) e 2 (*Me* = 3.86 mg/L) che si discostano dagli altri, i cui

Tab. 10 - Valori minimo, massimo e mediano per variabili caratterizzate da differenze significative nei vari periodi di campionamento come risultato dal test della mediana per a (significatività) compreso tra 0.05 e 0.01; N è il numero di campioni prelevati.

- Median, maximum and minimum values for variables with significant differences among the various sampling campaigns as derived by the median test for a between 0.01 and 0.05, N is the number of the collected samples.

Periodo di campionamento		Cl ⁻	SO ₄ ²⁻	NH ₄ ⁺
0	N	203	203	203
	<i>Minimo</i>	4.60	6.70	0.010
	<i>Massimo</i>	7500	1275	10.45
	<i>Mediana</i>	35.81	39.84	0.142
1	N	50	50	50
	<i>Minimo</i>	3.00	6.70	0.013
	<i>Massimo</i>	1874	650	1.64
	<i>Mediana</i>	19.09	33.75	0.18
2	N	50	50	50
	<i>Minimo</i>	7.8	7.5	0.026
	<i>Massimo</i>	2625	675	3.68
	<i>Mediana</i>	32.63	46.25	0.12
3	N	48	48	48
	<i>Minimo</i>	8.20	10.0	0.039
	<i>Massimo</i>	44000	8500	3.01
	<i>Mediana</i>	56.50	45.06	0.32
Totale	N	351	351	351
	<i>Minimo</i>	3.0	6.7	0.010
	<i>Massimo</i>	7500	1275	10.45
	<i>Mediana</i>	32.61	39.84	0.15

valori sono tendenzialmente più elevati.

Per quanto riguarda Na⁺, K⁺ e Br⁻ è ancora il monitoraggio 1 a presentare i valori inferiori (*Me* rispettivamente uguale ad 16.77, 2.25 e 0.04 mg/L), mentre nel caso del F⁻, risultano simili i valori del campionamento 0 (*Me* = 0.2 mg/L) e 1 (*Me* = 0.23), ma inferiori rispetto ai monitoraggi successivi. Infine, nel caso, del B risulta essere il monitoraggio 3 (*Me* = 0.034 mg/L) quello caratterizzato dal valore mediano inferiore.

Dall'analisi complessiva dei dati si può rilevare la presenza di un effetto di variabilità dovuto alla

stagionalità, rilevabile alla scala temporale considerata nel corso del presente studio. Tuttavia, la sensibilità delle variabili a tali variazioni non è sempre la stessa e, nel caso del Bacino dell'Arno, si possono classificare come variabili non sensibili TDS, HCO₃⁻+CO₃²⁻, Ca²⁺, Mg²⁺ e SiO₂, come variabili abbastanza sensibili Cl⁻, SO₄²⁻, NH₄⁺, ed infine come variabili molto sensibili pH, NO₂⁻, NO₃⁻, Na⁺, K⁺, F⁻, Br⁻ e B.

Nelle tavole I (b) - XVI (b) sono riportate, *dot map* utili per confrontare, dal punto di vista spaziale, il periodo di campionamento base nei confronti dei monitoraggi. Infatti, la scala cromatica per ogni variabile è relativa all'incremento della concentrazione, mentre le dimensioni del simbolo sono indicative del periodo di prelievo.

Come si può notare, in diverse situazioni si rilevano differenze tra il campionamento di maggio-agosto 2002 e quelli successivi in definite postazioni di campionamento, e in particolare lungo il medio-basso corso dell'Arno e nella zona meridionale.

È comunque importante ribadire che questi risultati sono da correlare con lo schema di campionamento e possono avere validità, quindi, per la scala temporale investigata. In altre parole, se effettuassimo controlli con una differente scala temporale di osservazione (settimanali, mensili, annuali, etc.) da quella scelta nel nostro caso, i risultati del test potrebbero cambiare, così come, probabilmente, anche i risultati relativi al comportamento delle specie considerate.

4.4. - GLI ELEMENTI IN TRACCIA

Nella Tavola 17 sono riportati i dati relativi ai valori di minimo, massimo e di mediana degli elementi Al, As, Ba, Cd, Cr, Cu, Fe, Li, Mn, Mo, Ni, P, Pb, Rb, Se, U, V, W, Y, Zn; i valori minori del limite di rilevabilità risultano essere in numero di 2 per As, 14 per Cd, 2 per Cr, 12 per Fe, 4 per Ni, 2 per Se, 2 per V e 4 per W. Nella Tavola 17 sono riportati anche i diagrammi a scatola per ognuna delle specie sopra elencate; al fine di confrontare la forma della distribuzione di frequenza dei vari elementi, i singoli box-plot sono stati costruiti dopo avere standardizzato i valori di ogni variabile in modo che ognuna avesse media uguale a zero e deviazione standard misurabile in unità di deviazioni standard. Tale obiettivo si raggiunge facilmente sottraendo ad ogni variabile la sua media e dividendo per la sua deviazione standard.

Dall'analisi dei diagrammi a scatola si rileva come Ba, U, Al, Fe, Mn siano caratterizzati da ampia variabilità nei valori, risultato che appare giustificabile considerando il comportamento

Tab. 11 - *Valori minimo, massimo e mediano per variabili caratterizzate da differenze significative nei vari periodi di campionamento come risultato dal test della mediana per $\alpha < 0.01$; N è il numero di campioni prelevati.*

- Median, maximum and minimum values for variables with significant differences among the various sampling campaigns as derived by the median test for $\alpha < 0.01$, N is the number of the collected samples.

Periodo di campionamento		pH	NO ₂ ⁻	NO ₃ ⁻	Na ⁺	K ⁺	F ⁻	Br ⁻	B
0	N	200	203	203	203	203	203	203	203
	Minimo	6.62	0.006	0.01	4.1	0.7	0.03	0.001	0.001
	Massimo	9.73	11.48	72.5	3980	153	2.00	27.50	6.28
	Mediana	8.08	0.12	3.41	36.50	3.18	0.20	0.10	0.12
1	N	41	50	50	50	50	50	50	50
	Minimo	7.47	0.010	0.60	4.2	0.6	0.040	0.005	0.001
	Massimo	8.33	0.315	85	1040	44.50	4.75	7.500	0.61
	Mediana	7.93	4.60x10 ⁻²	5.25	16.77	2.25	0.23	4.00x10 ⁻²	7.85x10 ⁻²
2	N	50	47	50	50	50	50	50	50
	Minimo	7.62	0.003	0.01	5.6	0.7	0.07	0.005	0.001
	Massimo	8.54	3.36	24	1487	58.00	7.75	9.250	2.24
	Mediana	8.01	0.16	0.3.87	034.20	3.32	0.28	7.50x10 ⁻²	4.92x10 ⁻²
3	N	48	48	48	48	48	48	48	48
	Minimo	6.97	0.007	0.35	7.2	1.0	0.065	0.005	0.001
	Massimo	8.80	1.66	150	2650	102	3.75	16.25	4.27
	Mediana	8.17	0.12	9.75	54.7	7.07	0.480	0.138	3.35x10 ⁻²
Totale	N	339	348	351	351	351	351	351	351
	Minimo	6.62	0.003	0.01	4.1	0.6	0.030	0.001	0.001
	Massimo	9.73	11.48	150	3980	153	7.75	27.50	6.28
	Mediana	8.05	9.84x10 ⁻²	4.27	33	2.94	0.230	0.10	9.10x10 ⁻²

geochimico degli elementi, variamente mobili in funzione delle condizioni di ossido-riduzione e del pH ed alla possibilità di formare complessi. In questo contesto, le loro distribuzioni di frequenza appaiono caratterizzate da forte asimmetria, con prevalenza di valori bassi ed estese code nelle regioni destre, influenzate dalla presenza di valori anomali elevati.

L'analisi della struttura complessiva di correlazione lineare tra gli elementi metallici, effettuata mediante il coefficiente non parametrico a causa sia della presenza di asimmetrie che di valori anomali, indica che As è l'elemento che maggiormente si correla con gli altri ed in particolare con V e Mo ma anche con P, Rb, Se, e W. Si rileva, inoltre, una debole correlazione tra Al, Zn e Cr e tra Mn, Mo e Li. Buone invece appaiono le correlazioni tra Rb, U e Li e tra Zn e P.