

LA STATISTICA PER VALUTARE LA POSSIBILE CONTAMINAZIONE

LA VALUTAZIONE DELLA PROBABILITÀ DI CONTAMINAZIONE DELL'ACQUA DESTINATA AL CONSUMO UMANO PUÒ ESSERE FATTA UTILIZZANDO TECNICHE STATISTICHE GIÀ DISPONIBILI. I RISULTATI DI UNO STUDIO SUL PARAMETRO CLORITO NELLA RETE IDRICA DI BOLOGNA MOSTRANO L'EFFICACIA DELLA TECNICA FAILURE MODE AND EFFECTS ANALYSIS (FMEA).

L'utilizzo di tecniche di tipo proattivo per la valutazione del rischio probabile di contaminazione dell'acqua destinata al consumo umano, è stato illustrato in un recente convegno organizzato dalla Regione Emilia-Romagna [1]; la tecnica utilizzata nota come *Failure Mode and Effects Analysis* (FMEA) è descritta nell'articolo a pagina 22 di questo numero di Ecoscienza [2].

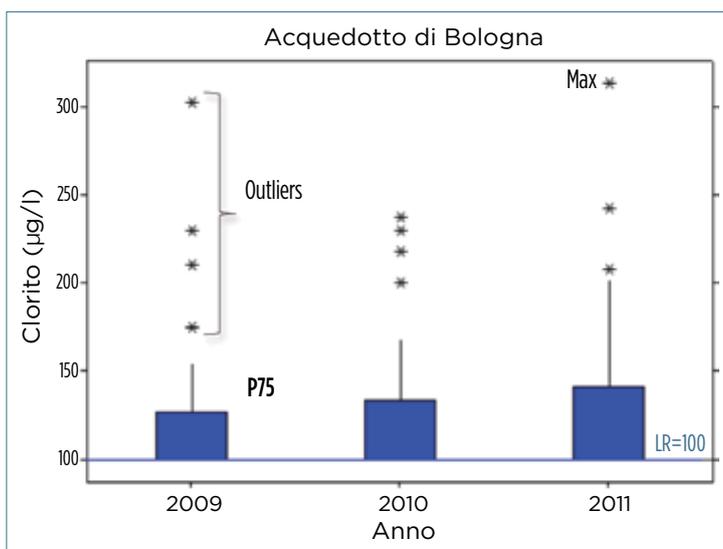
FMEA è uno strumento d'indagine impiegato per identificare, definire le priorità ed eliminare i potenziali errori nella progettazione di un prodotto o di un processo, prima che questi accadano [3]. FMEA assegna a ciascun evento avverso un *indice di priorità del rischio* (IPR) dato dal prodotto di tre indici: la *gravità* (G) dell'evento, la *probabilità* (P) che questo accada e la sua *rilevabilità* (R) [2, 3]. Gli indici G e P sono definiti preventivamente dalla conoscenza sul processo mentre l'indice R è legato alla possibilità di misurare l'evento avverso [4]. Nell'applicazione di FMEA al controllo delle acque potabili, l'indice R è stato ricavato dai risultati delle analisi eseguite nei laboratori di Arpa Emilia-Romagna [1].

In questa nota sono illustrati i metodi dell'analisi statistica (*box-plot*, istogramma, valore percentile, test non parametrici, distribuzione di Kaplan-Meier) utilizzati per analizzare un parametro di particolare interesse quale il *clorito*, rilevato nella rete idrica del comune di Bologna. L'analisi ha mostrato che questo parametro è molto stabile e ha permesso di individuare un indicatore riassuntivo – il 95° percentile dei valori disposti in ordine crescente – che è stato utilizzato per la stima del fattore R e per il calcolo dell'indice rischio probabile [2, 3]. I dati analizzati sono il risultato dei controlli eseguiti in un periodo di tre anni, che rappresentano le unità temporali di riferimento per l'analisi di stabilità.

Il clorito è stato scelto quale parametro di riferimento poiché riassume tutte le caratteristiche dei parametri controllati:

FIG. 1
CLORITO,
ACQUE POTABILI

Box-plot dei valori di clorito rilevati nell'acquedotto di Bologna, nel periodo 2009-2010.



Tab. 1 Valori riassuntivi del clorito rilevato nell'acquedotto di Bologna

Anno	Rilevazioni <i>n</i>	Undetected	Mediana	Percentile P95	0.95LCL	0.95UCL
2009	30	17	...	230	0.87	1.00
2010	31	15	100	237	0.87	1.00
2011	60	32	...	208	0.89	0.99

- limite massimo stabilito dal Dlgs 31/01
- distribuzione asimmetrica dei valori misurati con una coda di bassa frequenza alle alte concentrazioni e presenza di osservazioni inferiori al limite di rilevabilità strumentale LR.

L'analisi è stata eseguita con le tecniche statistiche sviluppate per i controlli sui dati ambientali [5] nei quali la presenza di valori minori del limite di rilevabilità (valori < LR o *undetected*) è molto frequente. Il *Box and Whiskers Plot* (*box-plot*) è la rappresentazione grafica più intuitiva di un data set di valori ordinati in modo crescente; il grafico è realizzato a partire dai valori di minimo, dal 25°, 50°, e 75° percentile e dal valore massimo. La posizione relativa del 25° (P25), 50° (P50 o mediana) e il 75° (P75) percentile, descrive il centro, la dispersione e l'asimmetria della distribuzione; i valori

anomali (*outliers*) se presenti sono indicati come singoli punti.

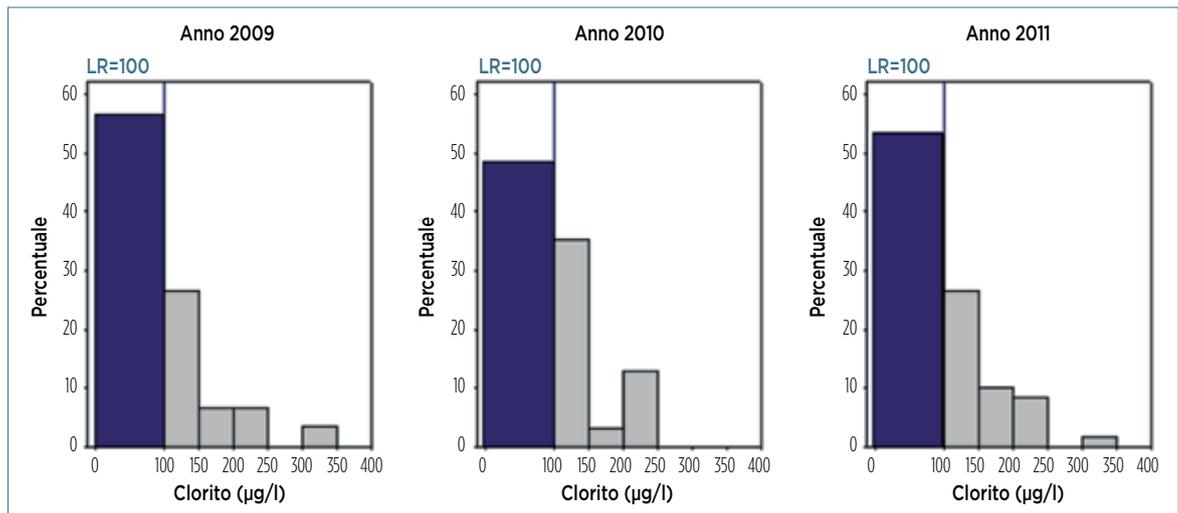
Se sono presenti valori *undetected* il *box-plot* è rappresentato solo per la parte di valori maggiori del limite di rilevabilità LR o *detected* [5].

Nel *box-plot* una linea orizzontale uguale a LR separa il *box-plot* in due regioni; i valori > LR sono rappresentati nel modo usuale, mentre i valori < LR non sono raffigurati dando origine a un *box-plot* troncato.

L'interpretazione del grafico è immediata: in *figura 1* i percentili P25 e P50 sono assenti e questo indica che, in ciascun anno preso in esame, il numero di valori < LR (*undetected*) è pari o superiore al 50% delle rilevazioni presenti nel campione. I valori dei percentili sono stati stimati con la tecnica di analisi non parametrica di Kaplan-Meier adattata allo studio di dati ambientali [5].

FIG. 2
CLORITO,
ACQUE POTABILI

Istogramma dei valori di clorito rilevati negli anni 2010-2012, nella rete idrica del comune di Bologna.



L'istogramma è la rappresentazione grafica utilizzata in alternativa al box-plot per evidenziare la presenza di valori *detected* e *undetected* nell'intervallo stabilito dal Dlgs 31/01.

Le rilevazioni eseguite all'interno della rete dell'acquedotto di Bologna provengono per il 70% dai punti di campionamento in uscita dalle centrali di potabilizzazione e le restanti da punti di rete [1]. In ciascun anno i valori *undetected* (<100 µg/l) sono circa il 50% e nel box-plot di figura 1 l'assenza della linea mediana, evidente negli anni 2009 e 2011, è una prova che questa proporzione è superiore al 50%.

Il test di Kruskal-Wallis afferma ($p\text{-value} = 0.87$) che non vi è evidenza sperimentale per affermare che i valori osservati nei tre anni siano significativamente diversi; in altri termini, la presenza di clorito è invariata. Il 95° percentile (P95) dei valori ordinati in modo crescente ha variazioni molto contenute e questa è un'indicazione che le condizioni fisico-chimiche che portano allo sviluppo di questa sostanza nei punti di prelievo presi in esame sono pressoché invariate.

In tabella 1 è riportato l'intervallo di confidenza, IC95%, calcolato a un livello di confidenza del 95%, come misura della stabilità di questo parametro nei tre periodi esaminati [5]. IC95% individua un intervallo di valori di percentile i cui estremi inferiore (95%LCL) e superiore (95%UCL) delimitano i valori di P95, se le n rilevazioni del clorito fossero ripetute in successione per cento (100) anni, supposto che la rete di Bologna mantenga invariate le sue caratteristiche. Gli intervalli sono molto simili a conferma che la concentrazione del clorito è molto stabile; in altri termini, se le condizioni fisico-chimiche nei punti di campionamento non variano il valore di

P95 rilevato nel 2010 (237 µg/l) potrebbe divenire l'ottantasettesimo percentile (P87) o nella migliore delle ipotesi il percentile P100; cioè, la concentrazione di clorito più elevata dell'intera rete. Per descrivere la distribuzione dei valori di clorito si è scelto un istogramma con classi di ampiezza 100 µg/l e 50 µg/l: la barra di colore blu mette in risalto i valori *undetected*, i restanti valori *detected* in barre di colore bianco sono distribuiti su una frazione dell'intervallo delimitato dal Dlgs 31/01 (700 µg/l).

Nei tre anni esaminati la distribuzione del clorito ha il medesimo profilo: per ciascun anno la proporzione dei valori <LR è compresa fra il 50% e il 60%, i valori >LR sono minori di 350 µg/l e la percentuale di valori nella classe di concentrazione maggiore è inferiore al 5%. Gli istogrammi hanno il medesimo profilo asimmetrico e il valore massimo si mantiene lontano dal limite di tossicità 700 µg/l.

I risultati delle analisi mostrano che gli strumenti statistici utilizzati possono descrivere efficacemente la distribuzione dei valori di un parametro molto importante come il clorito nella rete idrica del comune di Bologna alimentata da sorgenti profonde e superficiali, anche se sono presenti valori inferiori del limite di rilevanza LR, in percentuale rilevante. I grafici scelti, box-plot e istogramma, mostrano efficacemente come si distribuiscono i valori nell'intervallo stabilito dal Dlgs 31/01; infine, i test non parametrici sui percentili della distribuzione consentono di verificare la stabilità del parametro scelto. I risultati ottenuti mostrano inoltre che il 95° percentile è un indicatore efficace per rappresentare l'intera distribuzione del clorito e, nel medesimo tempo, mettere in risalto i valori di concentrazione più elevata causa di una possibile contaminazione.



In conclusione è ragionevole affermare che il percorso di analisi descritto in questa nota rappresenta uno strumento di monitoraggio dei parametri che caratterizzano la composizione dell'acqua potabile e che può essere utilizzato per la stima dell'indice di rilevanza R richiesto dalla FMEA [1, 2] e per il calcolo del rischio di una probabile contaminazione.

Claudio Bonifazzi¹, Samanta Morelli²,
Leonella Rossi²

1. Università di Ferrara
2. Arpa Emilia-Romagna

RIFERIMENTI BIBLIOGRAFICI

- [1] *Il controllo delle acque potabili: nuovi orizzonti*, convegno Regione Emilia-Romagna, Bologna 5 giugno 2013.
- [2] L. Rossi et al. "Analisi del rischio per le acque potabili. Come?", *Ecoscienza*, 3/2014.
- [3] D.H. Stamatis, *Failure mode and effect analysis*, Milwaukee, WI, ASQ Quality Press; 1995.
- [4] S. Rhee, K. Hishii. *Using cost based FMEA to enhance reliability and serviceability*, Adv. Eng. Inf., 17, (2003) 179-188.
- [5] D.H. Helsel. *Statistics for Censored Environmental Data, Using Minitab and R*, 2nd Edition. Wiley 2012.