Associazione Italiana di Acustica
47° Convegno Nazionale
24-28 maggio 2021

ORGANIZZAZIONE E ANALISI STATISTICA PRELIMINARE DEI DATI RACCOLTI NELLO STUDIO AIA-SNPA SUL RUMORE AMBIENTALE DURANTE L'EMERGENZA DA COVID-19

Giovanni Brambilla (1), Antonino Di Bella (2), Jacopo Fogola (3), Daniele Grasso (3)

1) CNR-INM Sez. Acustica e Sensoristica "O.M. Corbino", Roma, giovanni.brambilla@artov.inm.cnr.it
2) Dipartimento di Ingegneria Industriale – Università degli Studi di Padova, Padova, antonino.dibella@unipd.it
3) ARPA Piemonte, Torino, jacopo.fogola@arpa.piemonte.it, daniele.grasso@arpa.piemonte.it

SOMMARIO

Nell'ambito dell'accordo di collaborazione di ricerca tra l'AIA e il Sistema Nazionale per la Protezione dell'Ambiente (SNPA) è stato raccolto un consistente insieme di dati acustici rilevati in Italia durante l'emergenza da CoViD-19 e in anni precedenti. La presente comunicazione descrive i criteri adottati per l'organizzazione dei dati e delle informazioni disponibili, nonché le metodiche statistiche impiegate per una loro analisi preliminare.

1. Introduzione

L'accordo di collaborazione di ricerca tra l'AIA e il Sistema Nazionale per la Protezione dell'Ambiente (SNPA) è finalizzato all'analisi tecnico-scientifica dei dati acustici rilevati in Italia durante l'emergenza da CoViD-19, con molteplici obiettivi per valutare gli effetti sullo stato dell'ambiente, sui paesaggi sonori e sulla loro percezione.

Per la raccolta dei dati è stato inviato ai partecipanti un file Excel strutturato in vari fogli contenenti le istruzioni per la compilazione, le informazioni richieste su vari parametri caratterizzanti il rilevamento e il formato per il reporting dei livelli sonori. I rilievi sono diversificati a seconda che sia stato utilizzato un fonometro o uno smartphone con una apposita applicazione, come ad es. OpeNoise [1]. Per i dati fonometrici sono stati acquisiti i valori del livello equivalente orario L_{Aeqh} e la corrispondente data e ora, mentre per quelli da smartphone, più frequentemente di durata inferiore all'ora, i valori di L_{Aeq} erano cadenzati secondo il tempo di campionamento impostato dall'applicazione (1 s per OpeNoise).

Complessivamente sono pervenuti dati relativi a 255 siti, di cui 202 considerati validi e oggetto di successive elaborazioni, per un totale di oltre 226.000 campioni di L_{Aeqh}.

La presente comunicazione riguarda esclusivamente i dati fonometrici, aventi dimensione campionaria nettamente superiore a quella dei dati rilevati con smartphone, e ne descrive l'organizzazione e le metodiche statistiche impiegate per un'analisi preliminare.

2. Organizzazione dei dati

Il primo aspetto affrontato è stata l'organizzazione dei numerosi dati fonometrici di L_{Aeqh} raccolti, in relazione anche alle potenziali e molteplici analisi statistiche.

Si è iniziato, pertanto, a realizzare una matrice in formato Excel nella quale le righe corrispondessero alla data e all'intervallo orario e le colonne ai vari siti. Sono state aggiunte anche altre colonne riguardanti il giorno della settimana (Giorno, da 1=Lunedì a 7=Domenica), il tipo di giorno (colonna FSDL, 1-5=feriale, 5-6=fine settimana, 8=festivo) e un codice orario (da 0=intervallo orario 6:00-7:00 a 23=intervallo orario 5:00-6:00 del giorno successivo). Nel codice identificativo di ogni sito è stata aggiunta anche l'indicazione della sorgente sonora predominante riportata nella scheda dati ricevuta secondo la codifica elencata in Tabella 1.

Il file così ottenuto è stato importato in ambiente R [2] per l'analisi statistica e le rappresentazioni grafiche mediante script appositamente sviluppati, utilizzando anche librerie specifiche. Una prima analisi ha riguardato la dimensione campionaria dei dati di L_{Aeqh} nelle varie fasi temporali individuate nell'anno 2020 in relazione ai provvedimenti governativi adottati per contenere la diffusione della pandemia (Tab. 2). A fini comparativi la stessa suddivisione è stata adottata anche negli anni precedenti e successivi al 2020.

Tabella 1 – Codifica della sorgente indicata come predominante in ciascun sito.

Sorgente predominante	Codice	Sorgente predominante	Codice
Traffico stradale	S	Esercizi commerciali	C
Traffico ferroviario	F	Locali pubblici	LP
Traffico aereo	A	Vociare/schiamazzi	V
Suono naturale	N	Altro	AL
Attività produttive	P		

Tabella 2 – Suddivisione dell'anno solare in fasi.

Fase e limitazioni per il 2020	Inizio gg/mm	Fine gg/mm
A – Nessuna limitazione	1/1	25/2
L1 - Chiusura scuole e prime limitazioni ai movimenti	26/2	8/3
L2 - Ulteriori limitazioni ai movimenti e chiusura attività commerciali	9/3	22/3
L3 - Chiusura attività produttive non essenziali o strategiche	23/3	3/5
F2a - Inizio fase 2	4/5	17/5
F2b - Ulteriori riaperture	18/5	2/6
F2c - Riapertura mobilità tra regioni	3/6	13/10
L4 - Tre DPCM per limitazioni di orari per locali (ore 18) e coprifuoco ore 23-05	14/10	5/11
L5 - Istituzione zone gialle-arancioni-rosse	6/11	31/12

Un esempio di rappresentazione grafica dei dati disponibili nel singolo sito è riportato nella Figura 1, mentre nella Figura 2 è riportata la successione temporale dei dati disponibili per singola fase (nello specifico la fase F2a 2020) nei vari siti, con indicazione dei livelli L_{Aeqh} secondo una scala cromatica. È disponibile, inoltre, un grafico di maggiore dettaglio per ogni sito e fase temporale con la successione dei valori di L_{Aeqh} (Fig. 3).

Si sottolinea che le differenti modalità di rilevamento (ad es. distanza sorgente-ricettore) nei vari siti possono inficiare il confronto diretto dei livelli L_{Aeqh} tra i diversi siti, mentre è valido il confronto di tali livelli nello stesso sito tra le varie fasi temporali. Quest'ultimo quantifica il cambiamento degli ambienti sonori durante le restrizioni emanate per l'emergenza da CoViD-19 e le differenze risultanti sono confrontabili tra i vari siti.

Associazione Italiana di Acustica, 2021
ISBN: 978-88-88942-61-2

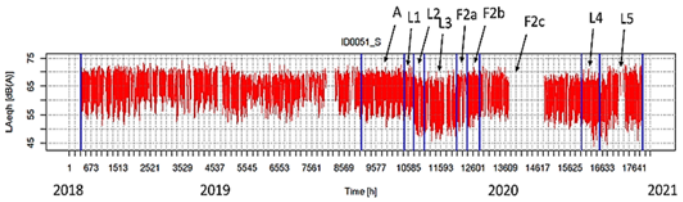

Figura 1 – Andamento dei livelli L_{Aeqh} nelle varie fasi temporali per singolo sito.

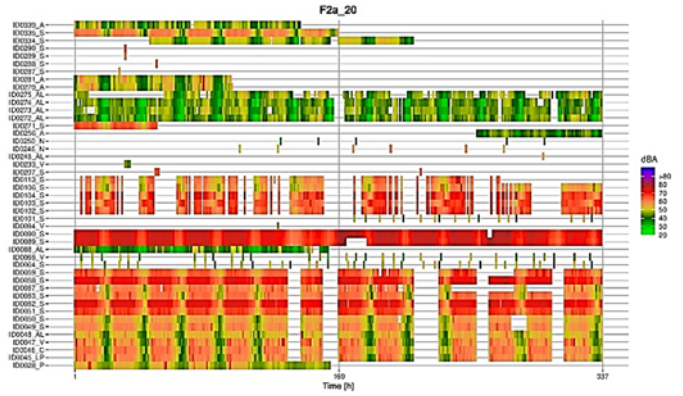

Figura 2 – Dati L_{Aeqh} disponibili per singola fase (F2a 2020) nei vari siti e loro rappresentazione su scala cromatica.

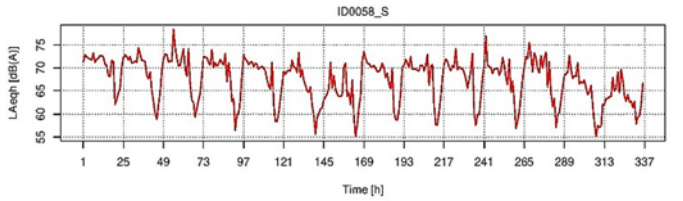

Figura 3 – Andamento dei livelli L_{Aeqh} per singolo sito (ID0058_S) e fase temporale (L2 2020 in figura).

3. Analisi statistiche descrittive preliminari

Per ciascuna fase temporale e singolo sito si è proceduto alla distinzione tra tempo di riferimento diurno TR_d (ore 6-22) e notturno TR_n (ore 22-6) e sui dati di L_{Aeqh} così diversificati sono stati calcolati i parametri statistici elencati nella Tabella 3.

Tabella 3 – Parametri statistici determinati per ogni fase temporale sulla serie di dati L_{Aeqh} per singolo sito e tempo di riferimento TR.

Numero dati L_{Aeqh} disponibili	Valore mediano L_{Aeqh} [dB(A)]
Dimensione campionaria rispetto alla fase [%]	Scarto s dei valori L_{Aeqh} [dB(A)]
L_{Aeqh} massimo [dB(A)]	Coefficiente di variazione [%]
L_{Aeqh} minimo [dB(A)]	Media aritmetica L_{Aeqh} troncata al 5% [dB(A)]
Media logaritmica L_{Aeqh} [dB(A)]	

Sono stati determinati anche per i TR_d e TR_n i corrispondenti livelli L_{Aeqd} e L_{Aeqn} e il corrispondente numero di campioni L_{Aeqh} disponibili, senza procedere ad alcuna "spalmatura" ove il numero dei campioni risultasse inferiore a quello di TR (16 o 8 rispettivamente). Sui livelli L_{Aeqd} e L_{Aeqn} è stata calcolata la mediana e la media logaritmica sia senza distinzione del tipo di giorno (variabile FSDL in Figura 1), che diversificando i giorni lavorativi (codice FSDL 1-5) dal fine settimana e giorni festivi (codice FSDL 6-8). Tutti i dati e i parametri calcolati sono stati esportati in un file Excel per agevolare la condivisione dei dati e dei risultati tra i partecipanti al gruppo di lavoro sull'analisi dati.

Un aspetto emerso nell'analisi delle serie temporali di L_{Aeqh} e meritevole di approfondimenti ha riguardato l'individuazione di dati "anomali", diversi da quelli indicati dal referente dei dati stessi solitamente in corrispondenza di condizioni meteo avverse o di eventi sonori atipici al contesto acustico. L'individuazione degli "outliers" è un argomento molto studiato in statistica con numerose proposte di tecniche numeriche di riconoscimento [3].

Per i dati L_{Aeqh} oggetto di analisi si è deciso di procedere con un duplice approccio sequenziale. La prima procedura, di tipo automatico, individua per ogni serie temporale in ciascun sito e fase i valori di L_{Aeqh} superiori a 3 volte lo scarto tipo. A rigore, questa procedura richiede l'ipotesi che i dati L_{Aeqh} siano distribuiti secondo una gaussiana. Il test di Shapiro-Wilk [4] ha evidenziato, come peraltro atteso, che nella stragrande maggioranza dei casi i valori di L_{Aeqh} non risultano distribuiti normalmente. Ciononostante, questo criterio è stato applicato per una prima discriminazione automatica dei potenziali dati "anomali". A questa fase segue un'analisi più approfondita e selettiva dei valori così individuati, che tiene conto del contesto acustico, della ricorrenza del dato in giorni precedenti e seguenti e così via. Al momento questa seconda analisi di dettaglio è in corso su un sottoinsieme di 28 siti caratterizzati da una cospicua serie di campioni di L_{Aeqh}, ricomprendente almeno la fase precedente la pandemia (fase A-ante) e quella di lockdown stretto occorsa nella primavera 2020 (L3). In questa analisi può risultare un utile riferimento il box plot dei valori di L_{Aeqh} per le singole ore, riportato a titolo esemplificativo nella Figura 4 per il sito ID0103_S nella fase temporale L3 2020 senza distinzione tra i giorni della settimana (variabile FSDL non considerata).

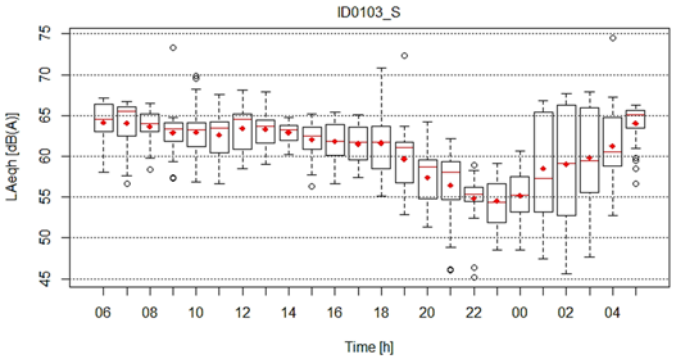

Figura 4 – Box plot per singola ora dei valori L_{Aeqh} per il sito ID0103_S nella fase temporale L3 2020 e senza distinzione tra i giorni della settimana.

4. Conclusioni

L'ingente insieme dei dati raccolti ha richiesto un impegno significativo per la loro organizzazione che è stata strutturata in funzione delle potenziali analisi statistiche. Per queste ultime sono stati utilizzati vari script appositamente sviluppati in ambiente R, impiegando anche varie librerie. Dati e risultati sono stati esportati anche in files Excel per agevolare la loro condivisione tra i partecipanti al gruppo di lavoro sull'analisi dati. Ulteriori analisi sono in corso, ad esempio in termini di L_{den} e L_{night}, nonché un approfondimento sulla individuazione dei valori anomali di L_{Aeqh} e un'analisi di sensibilità della loro influenza su valori aggregati (mediana e media logaritmica).

Il patrimonio di dati raccolti è di indubbio interesse e le molteplici modalità analitiche, tra le quali la comparazione tra le varie fasi temporali antecedenti, durante il lockdown e successive, potranno dare indicazioni interessanti sulle variazioni degli ambienti sonori e sulla loro percezione.

Bibliografia

[1] Masera et al., Realizzazione di un sistema di monitoraggio del rumore a basso costo attraverso la nuova app Android "OpeNoise", Rivista Italiana di Acustica, **40** (2016), pp. 48-58

[2] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

[3] Ranga Suri N.N.R., Narasimha Murty N., Athithan G., *Outlier Detection: Techniques and Applications. A Data Mining Perspective*. Springer Nature Switzerland AG, 2019

[4] Shapiro S.S., Wilk M.B., *An analysis of variance test for normality (complete samples)*, Biometrika, **52** (1965), pp. 591–611