

# Report Progetto Dati Funzionali

Matteo Ceola, Paolo Magagnato, Marco Piccolo e Pietro Stangherlin

## Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Obbiettivi</b>	<b>1</b>
<b>3</b>	<b>Dati</b>	<b>2</b>
<b>4</b>	<b>Operazioni preventive</b>	<b>2</b>
4.1	Rappresentazione funzionale . . . . .	2
4.1.1	Spline penalizzate e vincolate . . . . .	2
4.1.2	Risultati . . . . .	3
<b>5</b>	<b>Medie funzionali</b>	<b>5</b>
<b>6</b>	<b>PCA funzionale</b>	<b>5</b>
<b>7</b>	<b>ANOVA funzionale</b>	<b>5</b>
<b>8</b>	<b>Modello funzione su funzione</b>	<b>5</b>
<b>9</b>	<b>Conclusioni</b>	<b>5</b>

## 1 Introduzione

## 2 Obbiettivi

- Analisi esplorative funzionali
- ANOVA funzionale: confronto tra spettri di frequenze per diverse specie di uccelli
- Modello funzione su funzione: si è interessati a valutare se esistano delle relazioni tra ciascun suono emesso ed il suono precedente

### 3 Dati

I dati considerati sono presenti sul portale [xeno-canto](#). Per ogni audio è disponibile la specie di uccello e le coordinate geografiche del rilevamento.

### 4 Operazioni preventive

- Passaggio al dominio della frequenza tramite spettro medio
- Normalizzazione delle ampiezze

#### 4.1 Rappresentazione funzionale

##### 4.1.1 Spline penalizzate e vincolate

Per le curve di Ampiezza in funzione della frequenza si è scelta una rappresentazione in base bspline di grado 3. Inizialmente si sono considerate due penalizzazioni: la prima sul numero di basi e la seconda sull'integrale del quadrato della derivata seconda (per un numero di basi abbastanza alto fissato), per entrambi i casi si è considerato come riferimento il parametro che minimizzasse il criterio di GCV. Tuttavia questi due criteri non permettono il rispetto dei vincoli: 1) di non negatività della curva 2) di ampiezza non superiore a 1 (a causa della normalizzazione).

Per ciascuno dei due criteri sopra menzionati si sono quindi introdotti i vincoli nel problema di ottimizzazione che può essere scritto come un programma di programmazione quadratica per cui sono disponibili delle routine. Il programma di ottimizzazione quadratica nella sua forma più generale è definito come

$$\min_b (-d^T b + 1/2 b^T D b) \text{ s.t. } A^T b \geq b_0 \quad (1)$$

Per una singola osservazione funzionale, per il lisciamento tramite scrittura in funzioni di base la funzione da minimizzare rispetto a  $b$  è

$$(y - \Phi b)^T (y - \Phi b) + \lambda b^T P b \quad (2)$$

dove  $y$  è il vettore dei punti osservati  $\Phi$  è la matrice delle funzioni di base valutate nei punti osservati del dominio della curva e  $b$  è il vettore dei coefficienti,  $P$  è una generica matrice di penalità e  $\lambda > 0$  indica l'entità della penalizzazione (per un criterio non penalizzato è sufficiente porre  $\lambda = 0$ ). Minimizzare Equazione 2 equivale a minimizzare

$$-y^T \Phi b + b^T \frac{1}{2} (\Phi^T \Phi + \lambda P) b \quad (3)$$

Da cui  $d = y^T \Phi$  e  $D = \Phi^T \Phi + \lambda P$ . Usando la definizione di scrittura in basi il vincolo è  $0 \leq \phi^T(f) b \leq 1 \quad \forall f$ , in pratica si discretizza  $\phi(f_j)$  per  $j = 1, \dots, J$ .

Sia  $\Phi_J$  la matrice di funzioni di base valuate su griglia discretizzata: deve valere  $\Phi_J b \leq \mathbb{1}$  o equivalentemente  $-\Phi_J b \geq -\mathbb{1}$  dove con questa scrittura si intende che la disuguaglianza deve valere per ogni elemento dei vettori. Similmente, per vincolo di positività si ha  $\Phi_J b \geq \mathbb{0}$ , tuttavia, poichè per costruzione le basi bspline sono sempre non negative è sufficiente imporre  $\mathbb{b} \geq \mathbb{0}$  con  $\mathbb{b}_J$  matrice identità. Combinando le due espressioni si ottiene

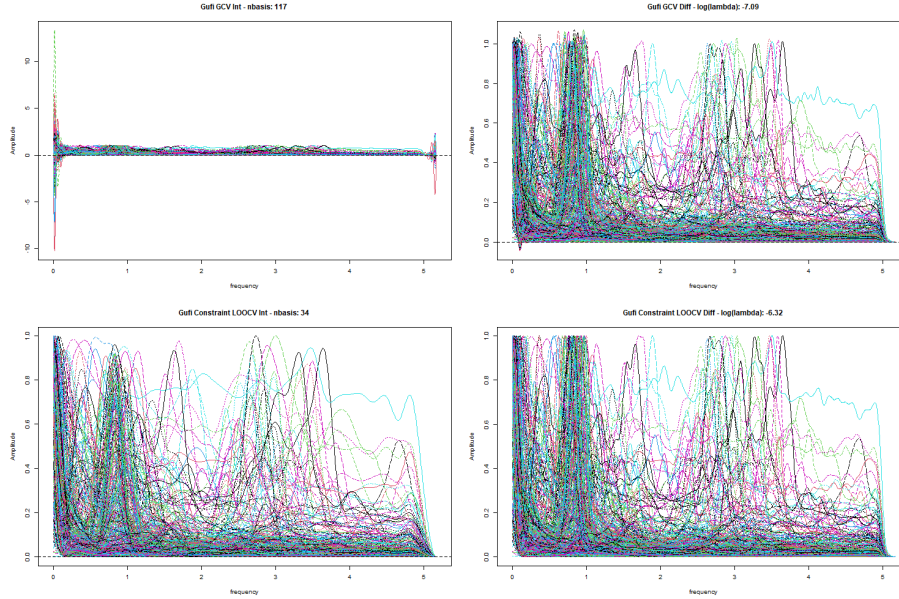
$$\begin{pmatrix} -\Phi \\ \mathbb{I} \end{pmatrix} b \geq \begin{pmatrix} -\mathbb{1} \\ \mathbb{0} \end{pmatrix}$$

chiaramente  $A = (-\Phi^T \quad \mathbb{I})$  e  $b_0 = (-\mathbb{1} \quad \mathbb{0})^T$ , è dunque conclusa la scrittura del problema vincolato come programma quadratico.

#### 4.1.2 Risultati

Introdurre i vincoli non dà luogo ad uno stimatore lineare in  $y$ , non è quindi possibile usare GCV come criterio per la selezione dei parametri di regolazione, si impiega invece una procedura di convalida incrociata “Leave One Out” (LOOCV). A titolo esemplificativo si riportano le curve relative ai gufi con i quattro metodi: GCV senza vincolo e LOOCV con vincolo; in @tab:representation-selection-df sono riportate le specifiche di ciascun metodo.

animal	constraint	penalty.type	min.error.parameter	domain.unique.points
falchi	FALSE	INT	9.00e+01	125
falchi	FALSE	DIFF	1.20e-06	125
falchi	TRUE	INT	4.80e+01	125
falchi	TRUE	DIFF	2.10e-06	125
gufi	FALSE	INT	1.17e+02	121
gufi	FALSE	DIFF	1.00e-07	121
gufi	TRUE	INT	3.40e+01	121
gufi	TRUE	DIFF	5.00e-07	121
gabbiani	FALSE	INT	9.50e+01	121
gabbiani	FALSE	DIFF	1.00e-07	121
gabbiani	TRUE	INT	3.10e+01	121
gabbiani	TRUE	DIFF	1.10e-06	121



Esaminando @fig-gufi\_fits\_crit si evidenziano diverse problematiche:

- nel caso senza vincoli e con penalizzazione solo sul numero di basi (grafico in alto a sinistra) il criterio GCV seleziona un numero eccessivo di basi (vicino al massimo possibile) che induce dei gravi problemi di comportamento erratico agli estremi, una possibile soluzione è selezionare manualmente il numero di basi osservando criticamente sia l'andamento dell'errore di GCV sia le funzioni risultanti.
- nel caso senza vincoli e con penalizzazione sull'integrale della derivata seconda al quadrato (grafico in alto a destra), benchè il miglioramento rispetto al primo caso sia notevole, si nota come le funzioni non rispettino i vincoli: alcune funzioni sono minori di zero (per frequenze piccole) e maggiori di 1.
- i vincoli migliorano chiaramente la rappresentazione funzionale, tuttavia, il numero di basi che minimizza LOOCV (grafico in basso a sinistra) è probabilmente troppo piccolo in quanto alcune funzioni hanno dei picchi troppo bassi, anche qui si potrebbe pensare di aumentare il numero di basi; nell'ultimo caso (penalizzazione sull'integrale della derivata seconda al quadrato) (grafico in basso a destra) una possibile critica è che le funzioni non siano abbastanza lisce.

Risultati simili si hanno anche con le altre due specie. Alla luce dei commenti fatti si è scelto di impiegare i coefficienti ottenuti tramite imposizione di vincolo con penalità sull'integrale del quadrato della derivata seconda.

## 5 Medie funzionali

In Figura 1 si riportano le medie e le deviazioni standard funzionali per tutti i gruppi di ciascun animale. Le deviazioni standard sono circa dello stesso ordine delle medie per quasi tutti i gruppi e tutti i punti.

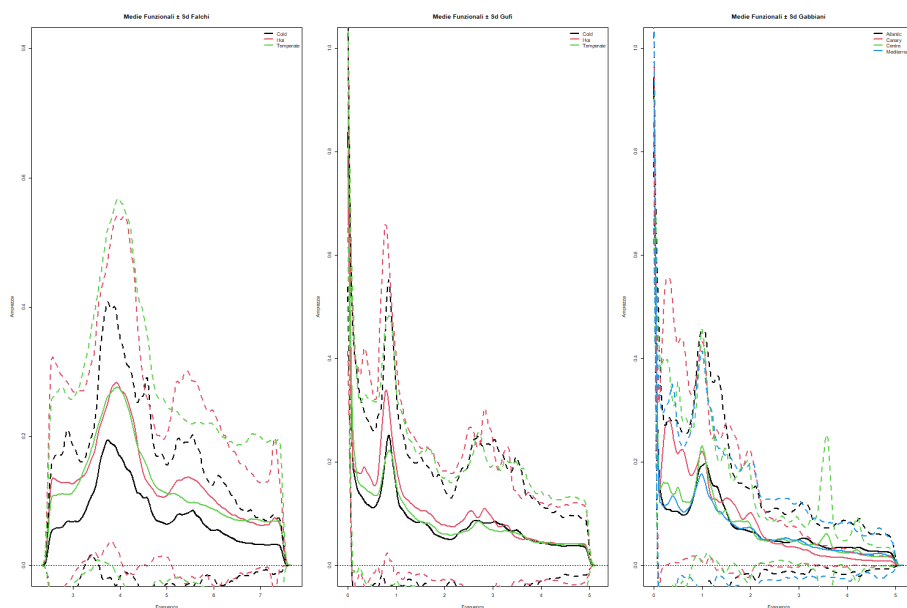


Figura 1: Medie Funzionali

## 6 PCA funzionale

Test Figura 4

## 7 ANOVA funzionale

## 8 Modello funzione su funzione

## 9 Conclusioni

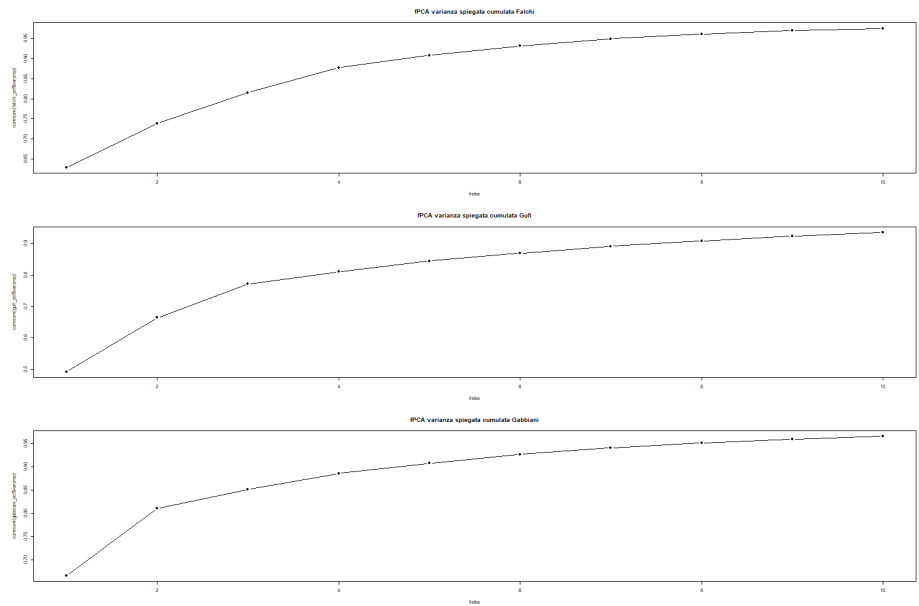


Figura 2: Varianza Spiegata Cumulata

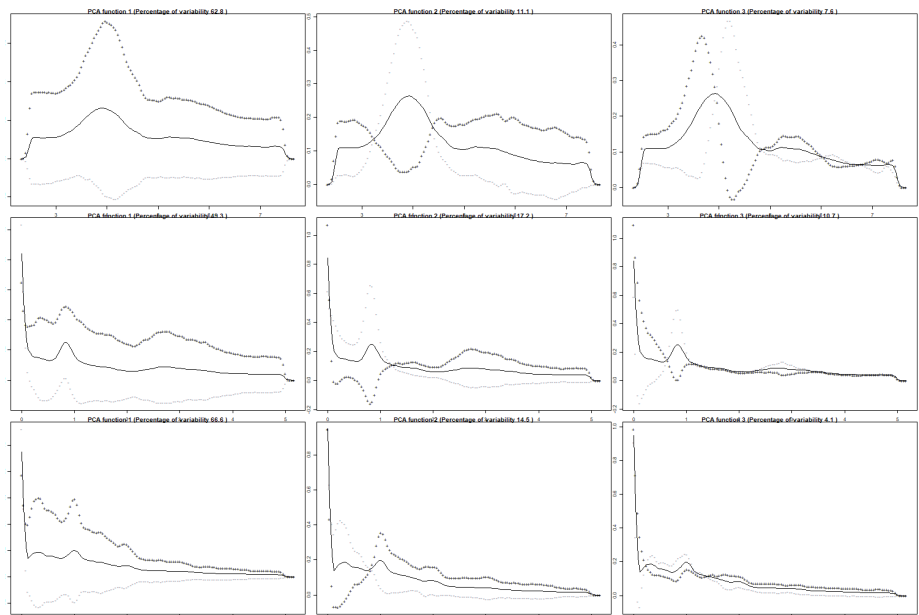


Figura 3: Armoniche

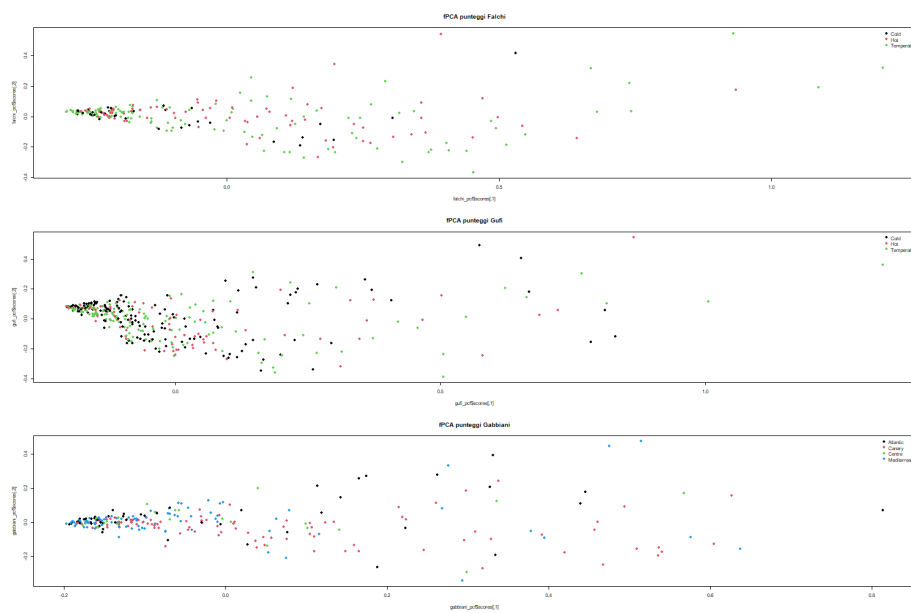


Figura 4: Punteggi