

# UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Modelli nulli per l’analisi della risposta di una rete  
neurale a perturbazioni esterne

Relatore

Dott. Michele Allegra

Laureando

Matteo Serafini

Anno Accademico 2023/2024



## **Sommario**

Uno dei problemi più comuni nelle neuroscienze applicate è quello di identificare quali unità, all'interno di una vasta e complessa rete neurale, rispondono a determinati stimoli esterni. A tale scopo, vengono tipicamente usati metodi di regressione (lineare o non lineare). L'ampiezza dei coefficienti di regressione, tuttavia, non è di per sé sufficiente a determinare in modo statisticamente rigoroso quali neuroni rispondono e quali no. Occorre, invece, fare considerazioni statistiche rigorose basate su modelli nulli che tengano conto di potenziali effetti spuri dovuti, ad esempio, alla non-indipendenza delle unità e all'autocorrelazione dei segnali. In questo lavoro di tesi, simuleremo una rete di neuroni, alcuni dei quali affetti da perturbazioni esterne, e proveremo a identificare i neuroni perturbati direttamente dai dati simulati, usando una regressione lineare e determinando la significatività dei risultati sulla base di diversi modelli nulli. Ciò permetterà di sondare l'efficacia e il rigore dei vari modelli analizzati.



# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Metodi per la simulazione e l'analisi della rete</b>	<b>3</b>
2.1	Simulazione della rete neurale . . . . .	3
2.1.1	Simulazione di un neurone isolato . . . . .	3
2.1.2	Simulazione di una rete di neuroni interconnessi . . . . .	5
2.1.3	Risposta ad una perturbazione esterna . . . . .	6
2.2	Regressione dei dati . . . . .	7
2.2.1	Modello descrittivo dell'effetto della perturbazione sul sistema . . . . .	7
2.2.2	Convoluzione con il segnale di calcio . . . . .	7
2.2.3	Test di ipotesi . . . . .	10
2.2.4	Correzione per comparazioni multiple . . . . .	10
2.3	Prewhitening del segnale . . . . .	12
2.4	Modelli nulli . . . . .	13
2.4.1	Pseudosessione . . . . .	15
2.4.2	Traslazioni lineari . . . . .	15
2.4.3	Traslazioni circolari . . . . .	15
<b>3</b>	<b>Analisi dei risultati della simulazione</b>	<b>17</b>
3.1	Confronto tra i metodi di analisi utilizzati . . . . .	17
3.1.1	Analisi dei segnali tramite test naïf . . . . .	18
3.1.2	Analisi dei segnali corretti tramite prewhitening . . . . .	19
3.1.3	Analisi tramite modelli nulli - senza correzione di prewhitening . . . . .	20
3.1.4	Analisi tramite modelli nulli - con correzione di prewhitening . . . . .	23
<b>4</b>	<b>Conclusioni</b>	<b>27</b>



# Capitolo 1

## Introduzione

Nelle neuroscienze, una questione centrale è capire come il cervello elabori determinati stimoli per produrre una risposta comportamentale. Il primo passo per affrontare tale quesito consiste generalmente nell'individuare quali componenti del sistema nervoso (aree cerebrali, colonne corticali, singoli neuroni) sono coinvolte nell'elaborazione dello stimolo. Una tipica classe di esperimenti si prefigge questo obiettivo, che può essere raggiunto identificando le componenti più attive in corrispondenza del (o immediatamente dopo) l'insorgere dello stimolo. In generale questo non è un compito di immediato assolvimento, perché l'attività cerebrale che viene registrata tramite gli esperimenti è rumorosa e tende a mescolarsi con un segnale di fondo dovuto al fatto che il cervello umano processa costantemente informazione, anche in assenza di una perturbazione esterna [6]. Ciò vale per registrazioni effettuate a tutte le scale, da quella macroscopica (risonanza magnetica funzionale, *fMRI*) a quella microscopica (registrazione elettrofisiologica o ottica dell'attività di singoli neuroni). Evidentemente questa classe di esperimenti non esaurisce lo spettro delle possibili indagini di un sistema complesso come una rete neurale biologica; tuttavia, a dispetto dell'apparente semplicità, permette di raccogliere una vasta mole di informazioni sulla struttura e il funzionamento del cervello.

In questa tesi ci si è concentrati su alcuni dei problemi che naturalmente emergono nell'analisi dei dati in questo tipo di esperimenti. Si descriveranno in dettaglio alcuni metodi che sono stati introdotti per cercare di affrontare tali criticità. Il raggio di applicabilità di tali metodi trascende largamente il ristretto ambito delle neuroscienze (sul quale ci si è focalizzati): la loro generalità, infatti, ne consente potenzialmente un'applicazione trasversale anche in campi molto diversi delle scienze applicate. Nel concreto, in questo lavoro si è costruita una rete di neuroni sulla base di un modello biologicamente affidabile [4] e si è introdotta nella dinamica di alcune unità scelte una perturbazione esterna. Si è interessati a determinare, solo studiando l'attività prodotta dalla rete, quali unità hanno risposto in maniera *significativa* allo stimolo aggiunto, ovvero quali neuroni hanno prodotto un'attività maggiore di quella che casualmente si potrebbe osservare in un sistema isolato. Normalmente, per rispondere a questo genere di domande, per capire quanto è probabile ottenere casualmente il risultato che si osserva, ci si affida ad un test statistico. Per diverse ragioni - e soprattutto per la non indipendenza tra le variabili in gioco - tali test falliscono nell'inferire correttamente se l'associazione tra stimolo e risposta è significativa o meno. Per fronteggiare questa difficoltà si è introdotta una classe di metodi, che generalmente va sotto il nome di **modelli nulli**, i quali rispondono precisamente alla domanda di partenza: "l'associazione che si osserva è più intensa di quella che si potrebbe osservare per caso?" Con i modelli nulli si creano delle copie equivalenti quasi in tutto al sistema di partenza, ma variando in maniera controllata alcuni parametri, così da "distruggere", qualora vi fosse, la relazione che sussiste fra le grandezze in gioco. Euristicamente, se l'associazione tra le variabili che si osserva in un sistema "scompigliato" è paragonabile a quella si trova nella configurazione di partenza si dovrà concludere che l'effetto dello stimolo non fosse realmente significativo. Se, al contrario, l'atto di scompigliare fa emergere associazioni *significativamente* inferiori a quella della configurazione di partenza si concluderà che lo stimolo ha di fatto avuto un effetto sull'attività cerebrale osservata.

Il lavoro svolto è stato dunque così suddiviso: nel Capitolo 2 si sono illustrati i modelli e i metodi tramite i quali è stata simulata una rete di neuroni biologici e in quali termini si è introdotta nel sistema una perturbazione esterna. Si sono poi esposti i metodi di preprocessamento dei dati ottenuti con la simulazione, al fine di renderli coerenti con quelli che normalmente si registrano in un esperimento reale. Infine si sono presentati tutti i metodi di analisi impiegati, dai più semplici ai più elaborati. Nel Capitolo 3 si sono invece presentati i risultati dell'analisi eseguita sui dati simulati per mezzo dei metodi previamente illustrati.



# Capitolo 2

## Metodi per la simulazione e l'analisi della rete

### 2.1 Simulazione della rete neurale

Si vuole simulare la dinamica di una rete di neuroni biologici, in prima istanza isolati e successivamente sotto l'effetto di una perturbazione esterna che agisce su un gruppo ridotto di unità. A tale scopo, è prima necessario specificare il modello matematico e i metodi numerici adottati per determinare l'evoluzione temporale di un singolo neurone. In seguito si procede costruendo un sistema accoppiato, cosicché la dinamica di ciascuna unità risulti influenzata anche da quella delle altre. Dapprima si considera una rete di neuroni isolata dall'esterno, affetta unicamente da un contributo stocastico dovuto all'effetto dell'ambiente in cui è naturalmente immersa. Infine, si introducono perturbazioni esterne che agiscono su alcuni neuroni della rete stessa.

#### 2.1.1 Simulazione di un neurone isolato

In letteratura esistono svariati modelli utili a descrivere la dinamica di un neurone biologico, di complessità e tipologia molto differenti. Per gli scopi di questo lavoro, risulta particolarmente valevole il modello di Izhikevich [4], il quale permette di ottenere un buon compromesso tra l'efficienza computazionale dei modelli *integrate-and-fire* e la plausibilità biologica di modelli più complessi di tipo *Hodgking-Huxley*.

Nel modello di Izhikevich la dinamica del neurone è controllata da un sistema differenziale nelle variabili  $u(t)$  e  $v(t)$  e nei quattro più uno parametri  $\mathbf{p} = (a, b, c, d)$ ,  $I = I(t)$ .

$$\begin{cases} \dot{v} = 0.04v^2 + 5v + 140 - u + I \\ \dot{u} = a(bv - u) \end{cases} \quad (2.1)$$

con la condizione aggiuntiva di *after-spike*, la quale aggiorna le variabili dinamiche come

$$\begin{cases} v \rightarrow c \\ u \rightarrow u + d \end{cases} \quad \text{se } v \geq v_{th} = 30 \text{ mV} \quad (2.2)$$

La variabile  $v(t)$  rappresenta il potenziale di membrana del neurone, mentre  $u(t)$  tiene conto delle correnti ioniche di attivazione e disattivazione e ha un effetto di feedback negativo su  $v$ . La variabile  $I(t)$  raccoglie l'insieme di tutti i segnali di corrente in ingresso al neurone e rappresenta il canale di interazione tra l'unità e il mondo esterno. A  $I(t)$  contribuisce un fondo stocastico, generalmente inteso come un contributo derivante da stimoli sporadici di origine talamica e indicato nel seguito come **input talamico**  $I^T$ . Questo termine, che risulterà sempre presente, costituisce l'alimentazione di base del neurone, senza il quale il potenziale di membrana tende a decadere verso un valore costante

entro un tempo caratteristico. All'input talamico si possono poi sommare altri termini in ingresso, dovuti a stimoli esterni o interazioni con altre unità ed è quello che verrà presentato nei paragrafi a seguire.

I quattro parametri liberi del sistema differenziale possono essere scelti in maniera tale da replicare il comportamento di diversi tipi di neuroni presenti nella regione corticale del cervello. Nelle simulazioni eseguite si sono utilizzati due tipi di neurone, a seconda del ruolo che svolgono nella rete:

- **Regular Spiking (RS)**
- **Low-Threshold Spiking (LTS)**

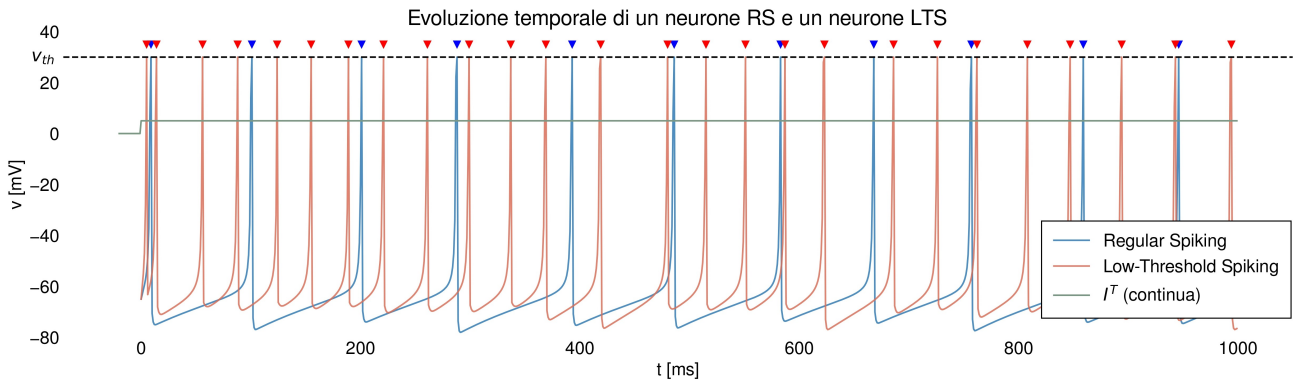
con parametri

$$p_{RS} = (0.02, 0.2, -65, 8) \quad p_{LTS} = (0.02, 0.25, -65, 2) \quad (2.3)$$

Per risolvere il sistema differenziale si sono utilizzati i metodi classici di risoluzione numerica delle equazioni differenziali, discretizzando l'asse temporale in intervalli di lunghezza  $dt$  e scrivendo il problema alle differenze finite tramite il metodo esplicito di Eulero con un doppio step da  $\frac{dt}{2}$ . Ciò permette una maggiore efficienza computazionale rispetto a metodi come il Runge-Kutta a quattro punti, senza però rinunciare eccessivamente alla precisione numerica del risultato. Formalmente

$$\begin{cases} v_{t+1}^* = v_t + \frac{dt}{2}(0.04v_t^2 + 5v_t + 140 - u_t + I_t) \\ v_{t+1} = v^* + \frac{dt}{2}(0.04v_t^2 + 5v_t + 140 - u_t + I_t) \\ u_{t+1} = a(bv_t - u_t) \end{cases} \quad (2.4)$$

Nel seguito si consideri sempre il passo temporale fissato a  $dt = 1$  ms. Per una simulazione preliminare si è scelto  $I^T(t) = 5$  mA, costante, al fine di valutare solo qualitativamente la correttezza dei risultati ottenuti. Si rimanda al paragrafo successivo per una trattazione più esaustiva dell'input talamico per come è stato descritto precedentemente. In figura è riportato il risultato di tale simulazione.



**Figura 2.1:** Evoluzione temporale di un neurone di tipo Regular Spiking (azzurro) e di un neurone di tipo Low-Threshold Spiking (rosso). In verde è riportato l'input talamico  $I^T$ , considerato per semplicità come un segnale in continua da 5 mA.

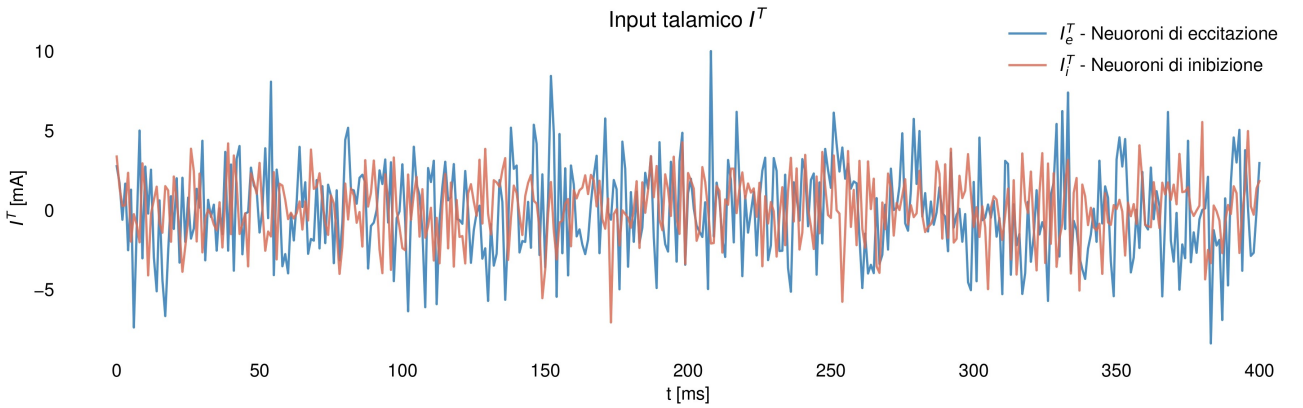
Così quindi si esplicita il meccanismo *integrate-and-fire*: il potenziale di membrana  $v$  evolve per effetto della corrente  $I$  che lo alimenta; quando  $v$  raggiunge un valore di soglia  $v_{th}$  viene emesso un **potenziale d'azione** (si parlerà in seguito di *spike*, perché quando tale condizione si verifica il potenziale di membrana risulta fortemente piccato all'istante  $t^*$  in cui il potenziale viene emesso). In questo modello il neurone è da considerarsi puntiforme, ovvero si trascura la sua struttura spaziale, ma in una cellula reale il potenziale d'azione si propaga lungo l'assone (il corpo centrale della cellula) e poi verso gli altri neuroni a lui collegati tramite le connessioni sinaptiche. Evidentemente per un neurone isolato l'emissione della spike non è così rilevante, giacché non influenza in alcun modo la dinamica del neurone stesso. Al contrario, nel paragrafo successivo l'effetto della spike è da tenere a mente per comprendere in che modo si intende modellizzare la connessione fra diverse le unità.

### 2.1.2 Simulazione di una rete di neuroni interconnessi

Verificato il corretto funzionamento di un neurone singolo si può dunque procedere a costruire la rete neurale. In questa simulazione si è considerato un sistema composto da  $N = 1000$  neuroni in cui ognuno di essi è da considerarsi connesso con i restanti  $N - 1$ . Delle  $N$  unità,  $N_e = 800$  sono di tipo eccitatorio e  $N_i = 200$  di tipo inibitorio, mantenendo così un rapporto di 4 a 1 tra di essi, coerentemente con l'anatomia della corteccia cerebrale dei mammiferi. I neuroni di eccitazione sono modellizzati tramite neuroni di tipo *RS*, mentre quelli di inibizione tramite neuroni di tipo *LTS*. Per ottenere un comportamento eterogeneo nella dinamica della rete, ai parametri 2.3 si aggiunge una componente stocastica, rimappandoli in

$$\mathbf{p}_{RS} = (0.02, 0.2, -65 + 15r^2, 8 - 6r^2) \quad \mathbf{p}_{LTS} = (0.02 + 0.08r, 0.25 - 0.05r, -65, 2)$$

$r \sim \mathcal{U}([0, 1])$  è una variabile casuale distribuita uniformemente in  $[0, 1]$  e il suo valore risulta scelto separatamente per ogni neurone. Sulla stessa linea di [4], ogni punto della serie temporale  $I^T(t)$ , campionata ogni  $dt = 1$  ms, è una variabile casuale distribuita normalmente e il suo valore è scelto separatamente per ogni neurone. Per i neuroni di eccitazione si ha che  $I_e^T(t_j) \sim \mathcal{N}(0, 5)$ , mentre per quelli di inibizione  $I_i^T(t_j) \sim \mathcal{N}(0, 2)$ . Si noti che queste distribuzioni sono tarate per riprodurre correttamente la dinamica dei neuroni solo qualora si mantenga lo stesso passo temporale fissato in precedenza: variando questo valore i parametri gaussiani vanno rimodulati se si vuole mantenere invariata la dinamica.

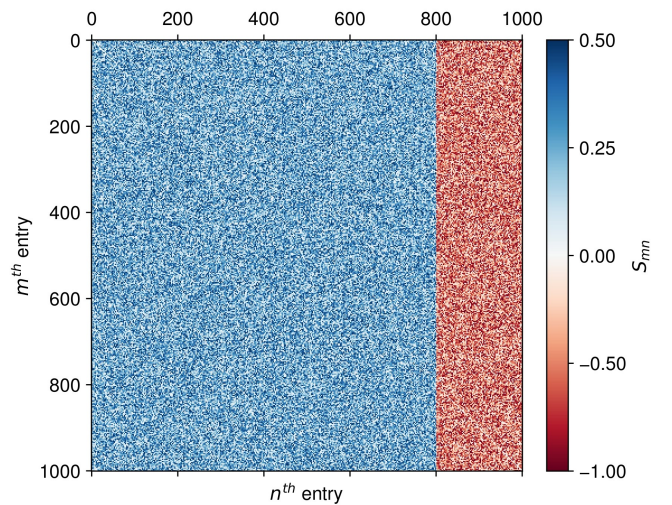


**Figura 2.2:** Esempio di corrente di input talamico cui sono soggetti i neuroni della rete. I punti della serie temporale sono distribuiti gaussianamente, con media 0 e deviazione standard che cambia in base al tipo di neurone.

Le connessioni sinaptiche tra i neuroni vengono modellizzate tramite una matrice  $\mathbf{S}$ , detta **matrice delle connessioni**, i cui coefficienti  $S_{mn}$  pesano l'effetto dell'  $n$ -esimo neurone sulla dinamica dell' $m$ -esimo. In particolare,  $S_{mn}$  è una variabile casuale distribuita in maniera uniforme nei due intervalli definiti in 2.5 e, in modulo, l'effetto dei neuroni di inibizione è mediamente più intenso di quelli di eccitazione:

$$\mathbf{S} : \begin{cases} S_{mn} \sim \mathcal{U}(D_e), & D_e = [0, 0.5) & \text{se } n \leq N_e \\ S_{mn} \sim \mathcal{U}(D_i), & D_i = (-1, 0] & \text{se } n > N_e \end{cases} \quad (2.5)$$

Nel concreto, se a un dato istante di tempo  $t^*$  si verifica che  $N^* \leq N$  neuroni (li si indicizza con  $n$ ) producono una *spike* - ovvero questi neuroni sono tali per cui  $v_n(t^*) \geq 30$  mV - ognuna delle  $N$  unità della rete risente di un contributo in

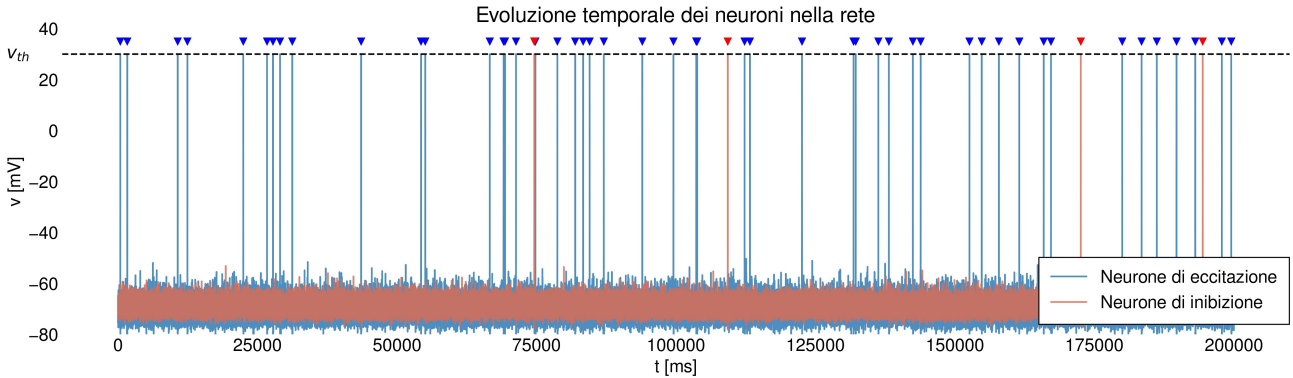


**Figura 2.3:** Rappresentazione grafica della matrice  $\mathbf{S}$  delle connessioni. La matrice è divisa verticalmente in due zone tra contributo relativo a neuroni di eccitazione e neuroni di inibizione.

corrente che si aggiunge a quello (casuale) talamico. Precisamente, per l' $m$ -esimo neurone tale contributo è dato dalla somma dei coefficienti  $S_{mn}$  al variare di  $n$  sugli  $N^*$  neuroni che hanno prodotto la spike. Sinteticamente

$$I_m(t^* + dt) = I_m^T(t^* + dt) + \sum_{n \neq m}^{N^*} S_{mn} = I_m^T(t^* + dt) + I_m^C(t^*) \quad \forall m \in \{1, \dots, N\}$$

La presenza di  $dt$  è da vedersi nell'ottica di un problema discretizzato: all'istante  $t^*$  il sistema evolve in maniera tale da produrre delle spike, le quali influenzano la dinamica della rete solo al passo temporale successivo.



**Figura 2.4:** Evoluzione temporale del neurone 1 (neurone di eccitazione), in azzurro, e del neurone 801 (neurone di inibizione), in rosso, parte di una rete di  $N = 1000$  neuroni.

### 2.1.3 Risposta ad una perturbazione esterna

Costruita la rete di neuroni su cui lavorare, si può ora introdurre una perturbazione esterna entro il sistema. Si è già menzionato il fatto che, nel sistema differenziale 2.1,  $I$  è il parametro che raccoglie tutti gli input (sotto forma di corrente) in ingresso al singolo neurone. In  $I$ , come già si è mostrato, convergono gli effetti dovuti alle spike prodotte degli altri neuroni connessi,  $I^C$ , e il fondo stocastico di origine talamica,  $I^T$ . A questo si può aggiungere una qualsiasi perturbazione che provenga dall'esterno,  $I^P = I^P(t)$ . La forma funzionale di  $I^P(t)$  è beninteso arbitraria e varia a seconda dello stimolo che si vuole esprimere. In questo lavoro si è costruita  $I^P(t)$  come una funzione a tratti che assume solo due valori e che presenta una struttura a blocchi: l'idea è quella di simulare l'effetto di uno stimolo binario (accesso-spento) sulla rete neurale. In particolare,  $I^P(t)$  è formato da  $n_P$  blocchi di ampiezza  $J$  e durata  $t_P$ , altrove si pone  $I^P(t) = 0$ . Il numero di blocchi varia in generale con la durata dell'osservazione al fine di mantenere una densità media di corrente indipendente dal tempo di simulazione. Data una simulazione di  $T$  millisecondi, gli  $n_P$  istanti  $t_i$  in cui ogni blocco comincia sono scelti casualmente nell'intervallo  $[0, T - t_P]$  secondo una distribuzione uniforme. Ovviamente gli  $n_P$  valori sono stati generati tenendo conto che

$$t_i \neq t_j + \eta t_P \quad \forall i \neq j \in \{1, \dots, n_P\}, \forall \eta \in [0, 1]$$

Sinteticamente, la perturbazione si può scrivere come

$$I^P(t) = \begin{cases} J, & \text{se } t \in D_{on} \\ 0, & \text{se } t \in D_{off} \equiv D_{on}^c \end{cases} \quad \text{con } D_{on} = \bigcup_i^{n_P} [t_i, t_i + t_P]$$

Ad ogni simulazione è stato generato uno stimolo  $I^P(t)$  nelle modalità specificate e sono stati estratti casualmente  $N_P$  neuroni tra gli  $N_e$  di eccitazione cui sottoporre tale perturbazione. Per questi neuroni quindi il parametro  $I$  consta di 3 termini:

$$I^* = I^T + I^C + I^P$$

mentre per i restanti  $N - N_P = 800$  il contributo alla corrente deriva soltanto dai primi due.

$$I = I^T + I^C$$

È bene chiarire infine che il segnale  $I^P(t)$  che viene generato casualmente in questo caso è unico per tutta la rete. Il senso di questo lavoro infatti è determinare quali neuroni rispondano ad uno stimolo macroscopico senza sapere *a priori* quali unità siano state di fatto veramente perturbate.

## 2.2 Regressione dei dati

### 2.2.1 Modello descrittivo dell'effetto della perturbazione sul sistema

Prima di introdurre il modello che esprime la relazione tra la perturbazione introdotta e la risposta del sistema è necessario costruire, a partire dal potenziale di membrana  $v_n(t)$ , la **serie degli spike**  $s_n(t)$ , una funzione definita come

$$s_n(t) = \begin{cases} 1 & \text{se } v_n(t) \geq v_{th} \\ 0 & \text{altrimenti} \end{cases}$$

ovvero una funzione che ad un dato istante  $t$  sia pari a uno e diversa da zero solo qualora in  $t$  venga emessa una spike. Questo passaggio è funzionale ad ottenere dei segnali da analizzare che siano coerenti con quelli rilevabili sperimentalmente. Infatti il potenziale di membrana non è una grandezza direttamente osservabile negli esperimenti, giacché l'attività cerebrale viene piuttosto codificata in termini della concentrazione di calcio rilasciato in seguito all'emissione di un potenziale d'azione. Da questo dunque l'esigenza di definire la serie degli spike: se un potenziale d'azione è emesso -  $s(t) = 1$  - il meccanismo del calcio si attiva, altrimenti -  $s(t) = 0$  - no. Al Paragrafo 2.2.2 verrà compiutamente spiegato come derivare a partire da  $s(t)$  il segnale di risposta da analizzare. Preliminarmente è necessario definire la relazione che sussiste tra la perturbazione esterna  $I^P(t)$  e la serie degli spike  $s(t)$ . A tale scopo si è deciso di introdurre un modello lineare della forma

$$s_n(t) = a_n I^P(t) + b_n + \epsilon_n(t) \quad \forall n \in \{1, \dots, N\} \quad (2.6)$$

Preso l' $n$ -esimo neurone  $a_n, b_n$  sono i coefficienti della regressione mentre in  $\epsilon_n(t)$  è raccolta la componente stocastica del segnale. La proposta di modelli non lineari più elaborati esula dagli scopi di questo lavoro e non è centrale nello studio dei metodi di inferenza statistica che verranno proposti. Peraltro, come si potrà constatare nel Capitolo 3, la descrizione che il modello proposto offre risulta già sufficientemente adeguata.

La regressione dei dati sulla base del modello 2.6 è stata ottenuta tramite il metodo dei minimi quadrati (*Ordinary Least Squares*, OLS). L'ipotesi di fondo su cui si basa tale metodo è che i punti su cui si effettua il fit siano variabili IID (Indipendenti e Identicamente Distribuite). Se sull'ipotesi di equa distribuzione si può anche soprassedere, altrettanto non si può dire circa la prescrizione d'indipendenza. Data la struttura del problema inevitabilmente i punti della serie temporale risultano correlati tra di loro: ciò risulta evidente se si considera come viene calcolato il valore di  $v_{t+1}$  a partire da  $v_t$  quando si propagano le equazioni dinamiche 2.1 tramite le relazioni alle differenze finite 2.4. In generale quindi è inevitabile che si abbia

$$\text{COV}[v_t, v_s] \neq 0 \iff \text{COV}[s_t, s_s] \neq 0 \iff \text{COV}[\epsilon_t, \epsilon_s] \neq 0,$$

almeno per valori prossimi di  $t$  e  $s$ .

### 2.2.2 Convoluzione con il segnale di calcio

Per ottenere il segnale osservato sperimentalmente a partire dall'informazione contenuta nella serie degli spike  $s_n(t)$  bisogna eseguire il prodotto di convoluzione tra  $s_n(t)$  e il segnale di risposta tipico

del calcio,  $f(t)$ , la cui forma analitica è data da

$$f(t; t_0, \tau_-, \tau_+) = \begin{cases} e^{\frac{t-t_0}{\tau_+}} & \text{se } 0 \leq t < t_0 \\ e^{-\frac{t-t_0}{\tau_-}} & \text{se } t \geq t_0 \end{cases} \quad (2.7)$$

con  $t_0 = 0.6$  s,  $\tau_+ = 1$  s,  $\tau_- = 4.8$  s. Si ricorda che, date due funzioni  $f(t), g(t)$  il loro prodotto di convoluzione  $(f * g)(t)$  è definito come

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)g(t - \tau)d\tau = \int_{-\infty}^{+\infty} g(\tau)f(t - \tau)d\tau \quad (2.8)$$

Alla serie degli spike si sostituirà allora la sua convoluzione con  $f(t)$ . Si definiscano formalmente le estensioni  $F, S_n, i^P$  ed  $E_n$  di  $f, s_n, I^P$  ed  $\epsilon_n$  a tutto  $\mathbb{R}$  come

$$S_n(t) = \theta(t)s_n(t) \quad F(t) = \theta(t)f(t) \quad i^P(t) = \theta(t)I^P(t), \quad E_n(t) = \theta(t)\epsilon_n(t)$$

dove  $\theta(t)$  è la *funzione gradino di Heaviside*. L'introduzione della distribuzione  $\theta$  è un passaggio formale utile a descrivere il fatto che la risposta del calcio è causale e la serie degli spike inizia convenzionalmente all'istante  $t = 0$ . Si definisce quindi il segnale  $y_n(t)$  come il prodotto di convoluzione

$$\begin{aligned} y_n(t) &= (F * S_n)(t) = \int_{-\infty}^{+\infty} F(\tau)S_n(t - \tau)d\tau \\ &= \int_{-\infty}^{+\infty} \theta(\tau)f(\tau)\theta(t - \tau)s_n(t - \tau)d\tau \\ &= \int_0^t f(\tau)s_n(t - \tau)d\tau \end{aligned}$$

Ora, recuperando il modello lineare introdotto per descrivere il rapporto fra  $s$  e  $I^P$ , si deve calcolare il prodotto di convoluzione con  $f(t)$  del secondo membro di 2.6. Si definisca la funzione

$$\mathbf{1}(t) = 1 \quad \forall t \in \mathbb{R}$$

Allora si deve valutare

$$y_n(t) = (F * S_n)(t) = a(F * i^P)(t) + b(F * \mathbf{1})(t) + (F * E_n)(t)$$

Si possono definire

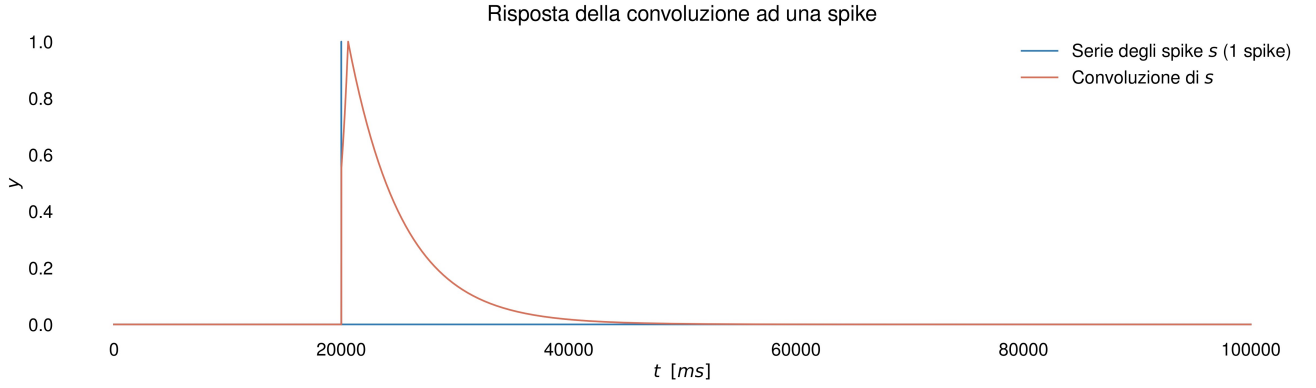
$$x(t) = (F * i^P)(t) = \int_{-\infty}^{+\infty} F(\tau)i^P(t - \tau)d\tau = \int_0^t f(\tau)I^P(t - \tau)d\tau \quad (2.9)$$

$$\epsilon'_n(t) = (F * E_n)(t) = \int_{-\infty}^{+\infty} F(\tau)E_n(t - \tau)d\tau = \int_0^t f(\tau)\epsilon_n(t - \tau)d\tau \quad (2.10)$$

Si mostra poi che il termine  $b(F * \mathbf{1})(t)$  resta costante sotto l'operazione di convoluzione. Infatti si ha

$$\begin{aligned} b(F * \mathbf{1})(t) &= b \int_{-\infty}^{+\infty} F(\tau)\mathbf{1}(t - \tau)d\tau \\ &= b \int_{-\infty}^{+\infty} \theta(t)f(\tau) \cdot 1d\tau \\ &= b \int_0^{+\infty} f(\tau)d\tau \\ &= b' \end{aligned}$$

Dalla seconda equazione di 2.9 emerge in maniera chiara come la convoluzione del segnale introduca un livello ulteriore di autocorrelazione entro il segnale, giacché la componente stocastica  $\epsilon'_n$  del segnale  $y$  all'istante  $t$  è una combinazione delle componenti stocastiche di  $s_n$  a tutti gli istanti  $t \in [0, t]$ .



**Figura 2.5:** Risposta della convoluzione ad una spike.

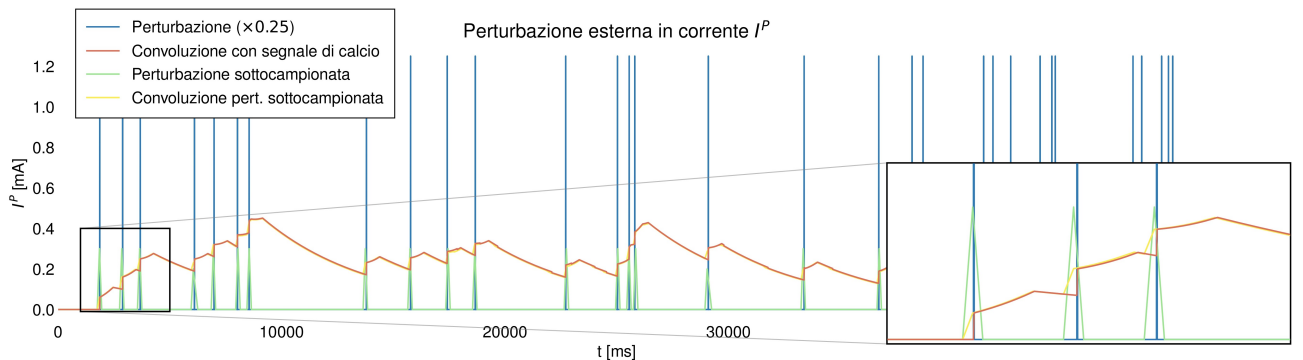
Prima di proseguire con la trattazione si rende esplicita, adottando da qui in avanti questa notazione, la natura discreta dei segnali con cui si sta lavorando. Quanto detto finora resta valido considerando le sostituzioni

$$g_n(t) \longrightarrow \mathbf{g}^{(n)} \quad \int d\tau \longrightarrow \sum_{\tau}$$

dove  $g_n(t)$  è una generica funzione del tempo relativa all' $n$ -esimo neurone e  $\mathbf{g}^{(n)}$  il vettore ottenuto campionando i valori di  $g_n$  ogni  $dt$  millisecondi. Si può riscrivere il modello 2.6, facendo cadere gli apici e sottintendendo che sono il risultato della convoluzione con  $f(t)$  come

$$y_t^{(n)} = a_n x_t + b_n + \epsilon_t^{(n)} \quad (2.11)$$

Si conclude questo paragrafo facendo menzione di un accorgimento pratico che si è dovuto adottare per implementare la convoluzione del segnale all'interno della simulazione. Per esigenze prettamente computazionali è stato necessario sottocampionare le due serie temporali  $\mathbf{s}^{(n)}$  e  $\mathbf{I}^P$ . In particolare, si è scelto un tempo di campionamento (*sampling time*,  $st$ ) e si sono rimappati i vettori  $\mathbf{s}^{(n)}$ ,  $\mathbf{I}^P$ , di dimensione  $N_t = \lfloor T/dt \rfloor$  ( $T$  essendo la durata della simulazione e  $dt$  il passo temporale con cui si è discretizzato l'asse temporale), nei vettori sottocampionati  $\tilde{\mathbf{s}}^{(n)}$ ,  $\tilde{\mathbf{I}}^P$  di dimensione  $N_t' = \lfloor T \cdot dt/st \rfloor$ . Si è indicata con  $\lfloor \cdot \rfloor$ , come di consueto, la funzione “parte intera”. L' $i$ -esima componente di  $\tilde{\mathbf{I}}^P$  è stata ottenuta mediando sugli  $st/dt$  valori di  $\mathbf{I}^P$  compresi fra gli istanti  $i \cdot st$  e  $(i+1) \cdot st$ . L' $i$ -esima componente di  $\tilde{\mathbf{s}}^{(n)}$  si è calcolata invece contando il numero di spike emesse tra gli istanti  $i \cdot st$  e  $(i+1) \cdot st$ . I vettori sottocampionati sono stati poi convoluti nelle modalità dichiarate sopra.



**Figura 2.6:** In azzurro, andamento temporale della perturbazione  $I^P$  introdotta nella dinamica di  $N_P$  neuroni scelti casualmente. In rosso, convoluzione di  $I^P$  con la risposta del calcio. In verde e in giallo il segnale sottocampionato e la convoluzione di quest'ultimo, rispettivamente. Il sottocampionamento è stato eseguito adottando un *sampling time* pari a 100 ms.

In figura 2.6 è riportato un esempio di perturbazione generata secondo le prescrizioni del Paragrafo 2.1.3 e del segnale di convoluzione. Si è anche riportato il segnale sottocampionato (ogni  $st = 100$  ms)

della perturbazione originale e la convoluzione ottenuta a partire da esso. Come si può notare le differenze sono trascurabili, soprattutto per tempi di simulazione molto più grandi di quelle di  $st$ , come in questo caso.

### 2.2.3 Test di ipotesi

Basandosi su un opportuno criterio, si può tentare di determinare, a partire dai soli coefficienti di regressione, se l'unità analizzata ha risposto o meno alla perturbazione introdotta. Questo è un approccio piuttosto *naïf* e, come verrà mostrato poi nel Capitolo 3, risulta di scarsa efficacia nell'inferire un risultato corretto. Nondimeno questo approccio risulta utile per fornire un metro di paragone per i modelli più raffinati che si intendono studiare. Pertanto, per rispondere in prima battuta alla domanda “La relazione tra stimolo e risposta è sufficientemente intensa da poter dire che il neurone è stato perturbato?” si è eseguito un test statistico. Precisamente si è utilizzato il  $t$ -test di Student sull'ipotesi nulla  $\mathcal{H}_0 : a = 0$ ; l'ipotesi alternativa, che si auspica di accettare solo per i neuroni perturbati, sarà data da  $\mathcal{H}_A : |a| > 0$ . Si ricorda che per il test di Student, nel caso di un modello lineare a due parametri come questo, la variabile  $t$  si calcola come

$$t = \frac{\hat{a} - 0}{S_a} \quad ndf = N_t - 2$$

ove  $\hat{a}$  è la miglior stima del coefficiente angolare e  $S_a$  l'errore della media associato al parametro.

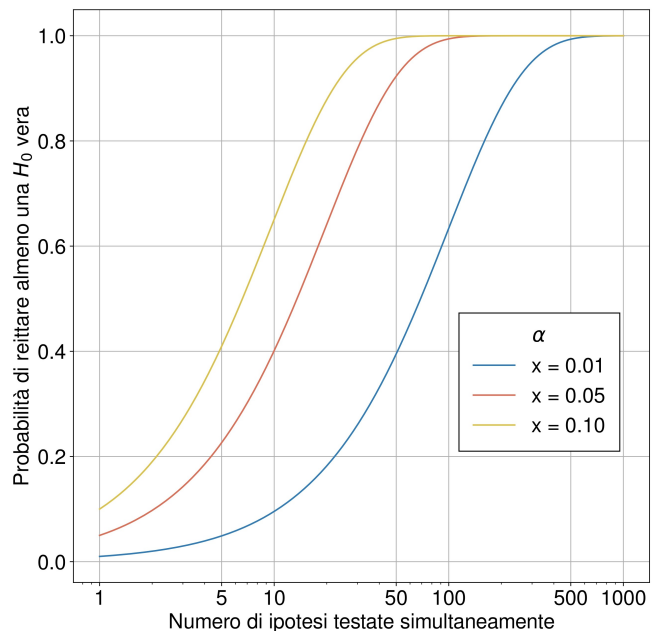
Una volta scelto un **livello di significatività**  $\alpha$ , a partire dalla distribuzione  $f_T(t)$  di Student si calcola il **p-value**  $p$  (la probabilità  $\mathcal{P}(T \geq t)$  di ottenere un valore di  $T$  che sia maggiore o uguale a quello misurato  $t$ ) e lo si confronta con la significatività scelta. Se  $p < \alpha$ , l'ipotesi nulla è da rigettare e si può ritenere che il neurone ha risposto alla perturbazione esterna.

### 2.2.4 Correzione per comparazioni multiple

Il livello di significatività  $\alpha$  che si sceglie rappresenta il tasso di errori di tipo I, o tasso di falsi positivi, che si è disposti a ottenere. Nel momento in cui si confronta più di un'ipotesi simultaneamente, tuttavia, la probabilità di commettere almeno un errore di tipo I cresce molto velocemente con il numero di ipotesi testate. Si consideri per esempio di volere testare  $N$  ipotesi nulle e per semplicità le si pensi tutte vere e indipendenti. La probabilità di rigettarne almeno una erroneamente è data da

$$\mathcal{P}(N_{\text{tipo I}} > 1) = 1 - \prod_{i=1}^N (1 - \alpha)$$

Se si considera come livello di significatività  $\alpha = 0.05$ , già per  $N = 100$  diventa di fatto inevitabile commettere almeno un errore di tipo I, considerato che  $\mathcal{P}_{N_{\text{tipo I}}}(N = 100) = 99.4\%$ . La stima della probabilità di errore diventa più complessa nel momento in cui le ipotesi nulle non siano tutte vere e vi sia correlazione tra di esse. Ciononostante l'esempio proposto è utile a capire la velocità con cui il tasso di errore cresce e la necessità di introdurre delle correzioni nei metodi di verifica delle ipotesi. Nel problema affrontato, sebbene non siano tutte vere, si testano contemporaneamente  $N = 1000$  ipotesi nulle ed è dunque necessario adottare le correzioni per confronti multipli che nel seguito verranno proposte.



**Figura 2.7:** Probabilità di commettere almeno un errore di tipo I testando  $N$  ipotesi simultaneamente.



Prima di fare ciò, tuttavia, è preferibile dare un quadro generale del problema e fornire alcuni concetti preliminari utili a comprendere la logica di funzionamento delle correzioni proposte. Si consideri dunque di voler testare  $N$  ipotesi simultaneamente e che di queste  $N_0 \leq N$  siano vere. In generale  $N_0$  è una variabile casuale non osservabile, ma nel problema specifico che si sta ora trattando è un valore noto e corrisponde al numero di neuroni imperturbati,  $N_0 = N - N_P = 980$ . Sia poi  $0 \leq R \leq N$  il numero di ipotesi nulle rigettate, che in qualsiasi caso risulta, evidentemente, una variabile osservabile. Qualora  $R > 0$ , si indichi con  $U \geq 0$  il numero di  $\mathcal{H}_0$  erroneamente rifiutate (si tratta per l'appunto del numero di errori di tipo I), che come  $N_0$  è in generale una variabile non osservabile.

I metodi proposti cercano tutti di controllare al meglio la variabile  $U$ , basandosi tuttavia su misure statistiche diverse. Una prima misura consiste nel considerare congiuntamente tutte le ipotesi  $\mathcal{H}_0$  come facenti parte di un'unica famiglia e ambire a non rigettarne erroneamente neppure una. In questo caso viene introdotta la misura statistica detta *FamilyWise Error Rate (FWER)*, definita come

$$FWER = \mathcal{P}(U > 0)$$

Generalmente la *FWER* è ritenuta una misura eccessivamente conservativa, specialmente nel caso in cui il numero di ipotesi testate sia molto alto. Solitamente è usata nel caso in cui il numero di  $\mathcal{H}_0$  sia dell'ordine delle decine. Sebbene in questo caso si stiano confrontando  $N = 1000$  ipotesi, sarà comunque utile adottare un metodo che si basi su questa misura come metro di paragone.

Una misura alternativa alla *FWER* è la *False Discovery Rate (FDR)*, definita come il valore di aspettazione del rapporto tra falsi positivi e reiezioni totali:

$$FDR = \begin{cases} \mathbb{E}\left(\frac{U}{R}\right) & \text{se } R > 0 \\ 0 & \text{se } R = 0 \end{cases}$$

Nel presentare i metodi di correzione proposti, si faranno le seguenti assunzioni. Siano  $\mathcal{H}_0^1, \dots, \mathcal{H}_0^N$  le  $N$  ipotesi testate e  $p_1 \dots p_N$  gli  $N$  p-value calcolati. Le ipotesi si devono considerare ordinate come  $\mathcal{H}_0^{(1)} \dots \mathcal{H}_0^{(N)}$  affinché  $p_{(1)} \leq \dots \leq p_{(N)}$ , ovvero in ordine crescente di p-value. Si indica come  $p'_i$  il nuovo valore dell' $i$ -esimo p-value ottenuto tramite la correzione per comparazioni multiple.

### Correzione di Hochberg

Di correzioni per comparazioni multiple che si basino sul controllo della misura *FWER* ve ne sono diverse, la più famosa essendo quella di Bonferroni. È tuttavia noto che questo sia generalmente un approccio piuttosto conservativo e si tende a preferire metodi più raffinati. In questo lavoro, pertanto, si è optato per la correzione di Hochberg. Dato un livello di significatività  $\alpha$  (corrispondente alla probabilità di commettere almeno un errore di seconda specie testando congiuntamente tutte le ipotesi), il metodo prevede di fissare il livello di significatività  $\alpha'_{(i)}$  della  $i$ -esima ipotesi testata come

$$\alpha'_{(i)} = \frac{\alpha}{N - i + 1} \longrightarrow p'_{(i)} = p_{(i)}(N - i + 1) \quad (2.12)$$

I test vengono dunque rieseguiti uno per volta a partire dall'ultima ipotesi. Non appena si ha che  $p_{(i)} < \alpha'_{(i)}$  il confronto si arresta e tutte le ipotesi nulle  $\mathcal{H}_0^{(j)}$  con  $j \in \{i, \dots, 1\}$  vengono rigettate con significatività  $\alpha$ . Evidentemente si può definire un p-value corretto  $p'_{(i)}$  come in 2.12 da confrontare con  $\alpha_{(i)}$ , ottenendo un risultato identico.

### Correzione di Benjamini-Yekutieli (BY)

La correzione di Benjamini-Yekutieli è un metodo di correzione per comparazioni multiple che, al contrario del precedente, controlla la misura di *FDR*. Detto  $q$  il limite superiore di *FDR* (per il quale si usano generalmente gli stessi valori convenzionali usati per  $\alpha$ ) si calcola il numero intero  $k$  come

$$k = \max_i \left\{ i \mid p_{(i)} \leq \frac{i}{N} \tilde{q} \right\} \quad \text{con } \tilde{q} = \frac{q}{\sum_{i=1}^N \frac{1}{i}} \quad (2.13)$$



con  $\epsilon_t = \mathcal{N}(0, \sigma)$ . Definendo  $\mathbf{e} = (e_1, \dots, e_{N'_t})$  si ha  $\mathbf{e} = \mathbf{B}\mathbf{e} + \boldsymbol{\epsilon}$  con

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \phi & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & \phi & 0 \end{pmatrix}$$

Poiché  $\mathbf{e} = (\mathbf{1} - \mathbf{B})^{-1}\boldsymbol{\epsilon} = \mathbf{A}^{-1}\boldsymbol{\epsilon}$ , si ha  $\mathbf{V} = \text{COV}[\mathbf{e}] = \mathbf{A}^{-1}\text{COV}[\boldsymbol{\epsilon}](\mathbf{A}^{-1})^T = \sigma^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^T$  (tenendo conto che gli  $\epsilon_t$  sono scorrelati e  $\text{COV}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{1}$ ). Quindi si arriva infine a

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2}\mathbf{A}^T\mathbf{A} \propto \mathbf{A}^T\mathbf{A}$$

Per un modello autoregressivo di ordine più alto,

$$e_t = \sum_{i=1}^p \phi_i e_{t-i} + \epsilon_t$$

con passaggi analoghi, e definendo la matrice  $\mathbf{B}$  come in Eq. (2.15), si ha ancora  $\mathbf{V}^{-1} = \frac{1}{\sigma^2}\mathbf{A}^T\mathbf{A}$

Per calcolare la radice quadrata di  $\mathbf{V}^{-1}$  è necessario procedere diagonalizzando la matrice. Detta  $\mathbf{D}$  la matrice diagonale composta dagli autovalori di  $\mathbf{V}^{-1}$  e  $\mathbf{U}$  la matrice unitaria che diagonalizza  $\mathbf{V}^{-1}$  come  $\mathbf{V}^{-1} = \mathbf{U}\mathbf{D}\mathbf{U}^\dagger$  infatti vale che

$$\mathbf{V}^{-1} = \mathbf{V}^{-1/2}\mathbf{V}^{-1/2} = \mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^\dagger\mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^\dagger = \mathbf{U}\mathbf{D}\mathbf{U}^\dagger$$

Sarà pertanto sufficiente calcolare  $\mathbf{D}^{-1/2} = \text{diag}(d_1^{1/2}, \dots, d_{N'_t}^{1/2})$  per ottenere tramite coniugazione la matrice di prewhitening  $\mathbf{W} = \mathbf{V}^{-1/2}$ . Ora, chiaramente la matrice va diagonalizzata tramite metodi numerici, ma il problema è complicato sostanzialmente da due fattori: il costo computazionale in prima battuta; l'utilizzo delle tecniche di manipolazione algebrica tramite matrici sparse, come conseguenza del primo fatto, in secondo luogo. Concretamente, per decomporre  $\mathbf{V}^{-1}$  si è sfruttato il metodo della decomposizione parziale ai valori singolari (*Partial Singular Value Decomposition, PSVD*). Evitando di entrare nei dettagli dell'algoritmo, nel concreto bisogna scegliere un intero  $k$  strettamente minore della taglia della matrice (in questo caso  $N'_t$ ) tanto più piccolo è questo intero, tanto peggio la decomposizione approssima la matrice di partenza. Per trovare un giusto compromesso tra precisione e velocità di esecuzione si è posto  $k = N'_t/2$ , avendo verificato che la correlazione tra la matrice di partenza e quella ottenuta tramite la decomposizione (l'indice di Pearson calcolato tra tutti gli elementi di una e tutti quelli dell'altra) era sistematicamente superiore al 95%.

4. In ultimo, si può dunque effettuare la regressione sul modello

$$\mathbf{W}^{(n)}\mathbf{y}^{(n)} = a\mathbf{W}^{(n)}\mathbf{x} + b\mathbf{W}^{(n)} + \mathbf{W}^{(n)}\boldsymbol{\epsilon}^{(n)} \quad (2.16)$$

il quale risulta ora ripulito dall'autocorrelazione che prima lo affliggeva.

## 2.4 Modelli nulli

Anticipando quanto verrà compiutamente mostrato nel Capitolo 3, dedicato ai risultati della simulazione prodotta e alla loro analisi, si può già dire che l'approccio *naïf* nella maggior parte dei casi fallisce nell'inferire correttamente un effetto significativo da parte della perturbazione su un dato neurone: pur tenendo conto della comparazione multipla di ipotesi, il tasso di fasi positivi che si registra risulta smodatamente elevato. Di fatto, una prima analisi di questo tipo fa emergere quelle che dallo statistico G.Yule vennero definite, in un articolo del 1926 [9], “*nonsense correlations*” - correlazioni insensate - tra le due serie temporali relative alla perturbazione e alla risposta neurale. Per ovviare

a questo problema si introduce ora una classe di metodi che vanno sotto il nome di **modelli nulli**. Propriamente - e in maniera molto generale - i modelli nulli rappresentano un copia alternativa del sistema analizzato, realizzata variando secondo una regola data alcuni parametri di questo e mantenendo fissi gli altri. Il confronto del sistema reale con un ensemble di “sistemi nulli” permette di disaccoppiare la dipendenza che intercorre tra alcune proprietà di interesse del sistema analizzato, di capire se e in che misura la presenza di una influenza l'emergere di un'altra e in generale di determinare se un certo comportamento osservato si manifesti solamente *per caso* oppure no. Per la loro flessibilità, specialmente nelle neuroscienze i modelli nulli si dimostrano essere strumenti di base per la ricerca in questo ambito, adattandosi efficacemente ai problemi per cui vengono impiegati [8]. Traducendo la definizione data nei termini del problema che si deve analizzare, le proprietà del sistema che si intendono mettere a confronto sono la perturbazione indotta dall'esterno e l'attività del sistema sotto risposta del calcio osservata. Si vuole determinare - lo si ripete solo per rendere chiara l'idea - quando l'insorgere di uno stimolo esterno influenza il normale tasso di spike della rete in una maniera non spiegabile solo tramite le normali fluttuazioni statistiche. Per costruire i modelli nulli si è mantenuta invariata la dinamica della rete, i parametri e i modelli che la definiscono quindi; soprattutto, è doveroso specificarlo, non si è cambiato in alcun modo lo stimolo esterno cui la rete è stata sottoposta. Ciò che è stato di volta in volta variato per costruire l'ensemble di confronto è il modo in cui la serie temporale della perturbazione si *abbinava* al segnale di risposta della rete. Qualora questa definizione non fosse sufficientemente esaustiva, lo diventerà una volta che si saranno descritte le regole tramite cui *riabbinare* le serie temporali.

Prima di presentare i diversi modelli adottati, è bene fornire alcune definizioni generali, che esulano dalla regola scelta in particolare, utili alla comprensione di quello che poi seguirà. Siano  $\mathbf{y}^{(n)}$  la serie temporale della risposta dell' $n$ -esimo neurone e  $\mathbf{x}^{(h)}$  la serie ottenuta per generare l' $h$ -esimo modello nullo a partire da  $\mathbf{x}$  secondo la regola fissata. Si denoterà con  $\mathbf{x}^{(0)}$  la serie non trasformata, vale a dire quella realmente osservata. Per poter creare una distribuzione di modelli nulli è necessario definire una funzione scalare detta **statistica di test**, che si indicherà successivamente come  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$ , che esprima il grado di associazione tra la serie della perturbazione e quella della risposta. In questo lavoro si è scelta come statistica di test la variabile  $\chi^2$  (considerando una incertezza unitaria sui punti  $y$ ): per ogni modello nullo  $h$  costruito sulla base della medesima regola si è ottenuto il vettore  $\mathbf{x}^{(h)}$  e con  $\mathbf{y}^{(n)}$  si è effettuato il fit lineare sul modello 2.11, calcolando la statistica di test  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$  come la somma dei residui al quadrato. Si è così ottenuta una distribuzione di probabilità per la variabile  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$ , cui nel seguito ci si riferirà come **distribuzione nulla**. Per ogni distribuzione nulla generata si è dunque calcolato un p-value come la “probabilità di ottenere un grado di associazione tra  $\mathbf{x}^{(h)}$  e  $\mathbf{y}^{(n)}$  maggiore di quello tra  $\mathbf{x}^{(h)}$  e  $\mathbf{y}^{(n)}$ ”. Quantitativamente, data la scelta fatta per  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$ , il p-value  $p$  è stato definito come

$$p = \mathcal{P}\left(V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)}) \leq V(\mathbf{x}^{(0)}, \mathbf{y}^{(n)})\right) \simeq \frac{1}{N_{M_0}} \mathcal{N}\left(V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)}) \leq V(\mathbf{x}^{(0)}, \mathbf{y}^{(n)})\right)$$

ove si è indicato con  $N_{M_0}$  il numero di modelli nulli generati e con  $\mathcal{N}(\cdot)$  il numero di modelli nulli che rispettino la condizione dell'argomento. Alla distribuzione nulla è d'uopo associare un'adeguata ipotesi nulla, che in questo caso è concettualmente la medesima già proposta alla Sezione 2.2.1, ma formulata secondo il paradigma dei modelli nulli:  $\mathcal{H}_0$ : “Il sistema osservato è estratto da un ensemble di modelli nulli”. Ovvero, in maniera più intuitiva, se il sistema realmente osservato è statisticamente compatibile con la distribuzione nulla, ovvero c'è un numero sufficientemente alto (per la significatività  $\alpha$  scelta) di modelli nulli in cui l'associazione (espressa dalla statistica di test  $V$ ) tra  $\mathbf{x}^{(h)}$  e  $\mathbf{y}^{(n)}$  è più forte, allora si conclude che la perturbazione non ha avuto un effetto maggiore di quello che si sarebbe potuto osservare casualmente. Se invece il sistema realmente osservato non è compatibile con la distribuzione creata, ovvero non si può considerare come parte di questo ensemble di modelli nulli, allora si deve concludere che l'effetto della perturbazione è effettivamente significativo. Anche in questo caso il livello di significatività è stato posto ad  $\alpha = 0.05$ . Nei prossimi paragrafi si illustrano le *regole* tramite le quali sono stati costruiti tre tipi di modelli nulli da impiegare nell'analisi.

### 2.4.1 Pseudosessione

Tramite il metodo della pseudosessione i modelli nulli vengono generati abbinando a  $\mathbf{y}^{(n)}$  una perturbazione  $\mathbf{x}^{(h)}$  aleatoria con le stesse caratteristiche statistiche della perturbazione realmente applicata (in questo caso, una perturbazione creata secondo le modalità espresse al paragrafo 2.1.3). Tale metodo richiede che le perturbazioni che agiscono sul sistema seguano una distribuzione di probabilità nota. Per questa ragione, in un'ottica più ampia in cui i modelli nulli vengano applicati a dati reali, questo metodo riscontra in un'applicazione, in generale, ridotta. Ciononostante, giacché la perturbazione cui è sottoposta la rete viene prodotta in maniera casuale secondo una distribuzione uniforme e delle regole note tale metodo può venir tranquillamente adoperato nell'analisi di questa simulazione.

### 2.4.2 Traslazioni lineari

Il metodo precedente, come si è fatto notare, richiede di includere in  $\mathcal{H}_0$  un'ipotesi forte e sperimentalmente non sempre verificabile, cioè il fatto che la serie temporale della variabile dipendente (in questo caso, la serie temporale delle perturbazioni) segua una distribuzione di probabilità nota. Per ampliare il campo di applicabilità dei modelli nulli impiegabili su dati reali, si può far cadere questa ipotesi a costo di introdurne due nuove. Si richiederà, in una formulazione generale di questo modello nullo, che almeno una delle due serie temporali (qui, in particolare,  $\mathbf{x}$ ) sia **stazionaria**, cioè che valga

$$\mathcal{P}(\mathbf{x}_{t_1}, \mathbf{x}_{t_2}, \dots, \mathbf{x}_{t_s}) = \mathcal{P}(\mathbf{x}_{t_1+\tau}, \mathbf{x}_{t_2+\tau}, \dots, \mathbf{x}_{t_s+\tau}) \quad \forall \tau, t_1, t_2, \dots, t_s$$

Sostanzialmente richiedere che il sistema sia stazionario equivale a supporre che sia invariante sotto traslazioni temporali e non esista un tempo assoluto di riferimento. Per come è stata generata la perturbazione, questa ipotesi è automaticamente verificata per costruzione. Accanto a ciò si dovrà anche richiedere che almeno una delle due serie sia **ergodica**, ovvero che valga l'*ipotesi di ergodicità*:

$$\mathbb{E}[\varphi(\mathbf{x}_{t_1}, \mathbf{x}_{t_2}, \dots, \mathbf{x}_{t_s})] = \lim_{T \rightarrow \infty} \frac{1}{2T+1} \sum_{\tau=-T}^T \varphi(\mathbf{x}_{t_1+\tau}, \mathbf{x}_{t_2+\tau}, \dots, \mathbf{x}_{t_s+\tau}) \quad \forall \varphi, t_1, t_2, \dots, t_s$$

In generale, mostrare che un sistema sia ergodico è cosa alquanto complicata. In questo caso, data ancora la costruzione del problema, è realistico assumere che sia verificata. Nel concreto, se il sistema è ergodico, sarà importante considerare serie sufficientemente lunghe da poter considerare verificato il limite in 2.4.2.

Ora, indicando come  $\mathbf{x}[t:s]$  tutti i punti di  $\mathbf{x}$  compresi fra gli istanti  $t$  ed  $s$ , sulla base delle due ipotesi richieste i modelli nulli per questo metodo si costruiscono confrontando  $\mathbf{y}^{(n)}[0:\sigma]$  con  $\mathbf{x}^{(h)} = \mathbf{x}[h:h+\sigma]$ , per  $h \in \{0, \dots, N'_t - \sigma\}$  e per un parametro  $\sigma$  scelto in partenza. In questo lavoro si è posto  $\sigma = N_t/2$ , dove, lo si ricorda,  $N'_t$  è il numero di punti della serie temporale. Si calcola quindi la statistica di test  $V(\mathbf{x}^{(0)}, \mathbf{y}^{(n)})$  e la si confronta con l'ensemble di statistiche di test  $\mathbb{V}_0 = \{V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)}) \mid h \in \{0, \dots, N_t - \sigma\}\}$  calcolate sui modelli nulli  $h$ .

### 2.4.3 Traslazioni circolari

Il metodo precedentemente presentato ha lo svantaggio di non sfruttare tutta l'informazione contenuta nei dati. Per superare questo limite, a costo di rendere più stringente l'ipotesi di stazionarietà, in questo metodo si considerano traslazioni circolari per una delle due serie temporali. In questo caso in particolare, per generare l' $h$ -esimo modello nullo, si confronta la serie  $\mathbf{y}^{(n)}$  con la serie traslata circolarmente  $\mathbf{x}^{(h)}$ . Precisamente,  $\mathbf{x}^{(h)}$  si ottiene rimappando i primi  $N'_t - h$  punti di  $\mathbf{x}$  in  $x_t \rightarrow x_{t+h}$  e gli ultimi  $h$  punti come  $x_t \rightarrow x_{N'_t-t+h}$ . Come prima dunque si confronta la statistica di test  $V(\mathbf{x}^{(0)}, \mathbf{y}^{(n)})$  ottenuta per la configurazione di partenza (senza traslazione) con l'ensemble  $\mathbb{V}_0 = \{V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)}) \mid h \in \{0, \dots, N_t - 1\}\}$ . L'impiego di questo metodo richiede, come detto, un'ipotesi di stazionarietà più stringente: la serie temporale che viene trasformata deve essere **ciclostazionaria**, ovvero invariante sotto traslazioni circolari. Per la costruzione del problema, ad ogni modo, tale ipotesi può nuovamente essere assunta come verificata.



# Capitolo 3

## Analisi dei risultati della simulazione

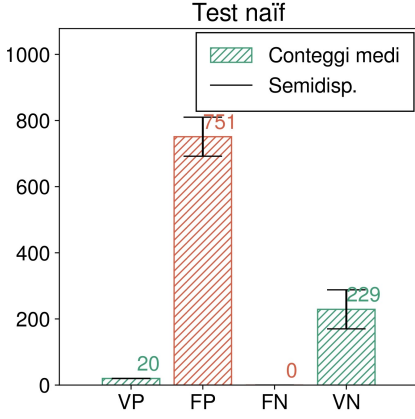
### 3.1 Confronto tra i metodi di analisi utilizzati

Nel capitolo precedente sono state definite le modalità tramite le quali costruire una rete di neuroni interconnessi, simularne l'evoluzione temporale e la risposta a una perturbazione esterna. Si sono quindi presentati gli strumenti statistici tramite i quali inferire, a partire solamente dal tasso di spike emesse, se un dato neurone ha risposto significativamente allo stimolo introdotto. Nel seguito verranno dunque esposti i risultati di questa analisi, introducendo in maniera progressiva le correzioni e gli strumenti di cui si è fatta menzione, per verificare di volta in volta la loro efficacia nell'inferire un risultato corretto. Si comincia quindi, per la configurazione più semplice, con quello che è stato chiamato gergalmente "test naïf", ovvero un fit lineare la cui significatività è dedotta tramite test di Student; in seconda battuta si introduce la correzione di prewhitening per la rimozione dell'autocorrelazione che affligge le serie temporali; in ultimo si applicano i modelli nulli proposti. Per tutti e tre i livelli di complessità si applica la correzione per comparazioni multiple.

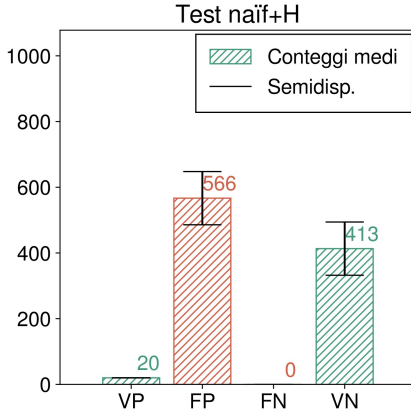
La simulazione prodotta ha una durata temporale di  $T = 300000$  ms, con un passo temporale fissato a  $dt = 1$  ms. I neuroni della rete restano, per qualsiasi configurazione della simulazione  $N_e + N_i = 800 + 200$ . Di questi se ne sono scelti  $N_p = 20$  tra quelli eccitatori e nella loro dinamica si è introdotta una perturbazione di intensità pari a  $\mathcal{J} = 5$  mA e formata da  $n_p = 100$  impulsi di corrente della durata di  $t_p = 5$  ms (mediamente un impulso ogni 3 s è dunque la densità scelta). Il sottocampionamento è stato effettuato con un tempo di campionamento pari a  $st = 100$  ms. Conseguentemente i punti delle serie temporale a disposizione per l'analisi sono  $N'_t = T/(dt \cdot st) = 3000$ : questo fatto sarà rilevante nei paragrafi a seguire quando si analizzeranno i modelli nulli. Fissati questi parametri si sono quindi eseguite 5 simulazioni, le quali differiscono l'una dall'altra per le proprietà del sistema lasciate libere di variare casualmente, vale a dire l'indice dei neuroni perturbati, gli istanti in cui gli impulsi della perturbazione vengono prodotti, i valori ad ogni istante dell'input talamico, gli della matrice delle connessioni, la componente stocastica dei parametri  $\mathbf{p}$  nel sistema differenziale 2.1. Nei grafici a seguire, assieme ai conteggi medi (troncati alla loro parte intera) è riportata una barra d'errore ad indicare la semidisperisione massima  $\Delta$  dei valori ottenuti nelle simulazioni. Nelle tabelle in allegato, invece, i conteggi medi (non troncati) sono riportati assieme ad un errore calcolato su  $\Delta$  tramite la statistica uniforme. Non è stato possibile realizzare simulazioni in numero e durata superiori a causa delle forti limitazioni nella capacità computazionale disponibile. Per fare un'analisi piuttosto qualitativa della complessità computazionale, si può constatare facilmente che il tempo di calcolo necessario ad evolvere il sistema dinamico scala linearmente con il tempo di simulazione  $T$  e con una potenza compresa fra 1 e 2 del numero di neuroni: la relazione è lineare qualora nessun neurone faccia una spike e quadratica qualora tutti i neuroni emettano una spike. Per quanto riguarda il processamento delle serie tramite l'algoritmo di prewhitening, invece, l'analisi è più complessa perché sono coinvolte operazioni diverse tra matrici. Per dare un'idea dell'ordine di grandezza, tuttavia, si pensi che la complessità del prodotto tra due matrici  $N \times N$  è dell'ordine di  $\mathcal{O}(N^3)$ , sebbene

possa essere ridotta con algoritmi più efficienti; la complessità dell'algoritmo *PSVD*, in questo caso, scala all'incirca come  $\mathcal{O}(N^3)$  [7]. Questa semplice analisi, per quanto rozza, evidenzia chiaramente la complessità algoritmica della simulazione, e la ragione dei limiti computazionali che si sono riscontrati.

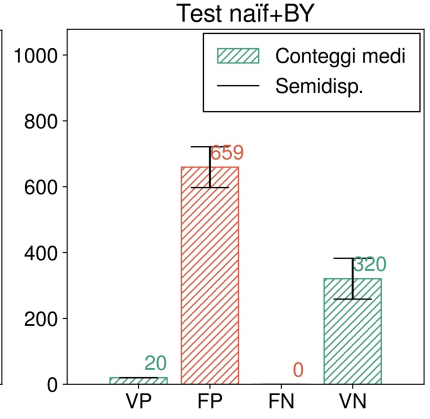
### 3.1.1 Analisi dei segnali tramite test naïf



**Figura 3.1:** Test naïf senza correzioni.



**Figura 3.2:** Test naïf con correzione di Hochberg.

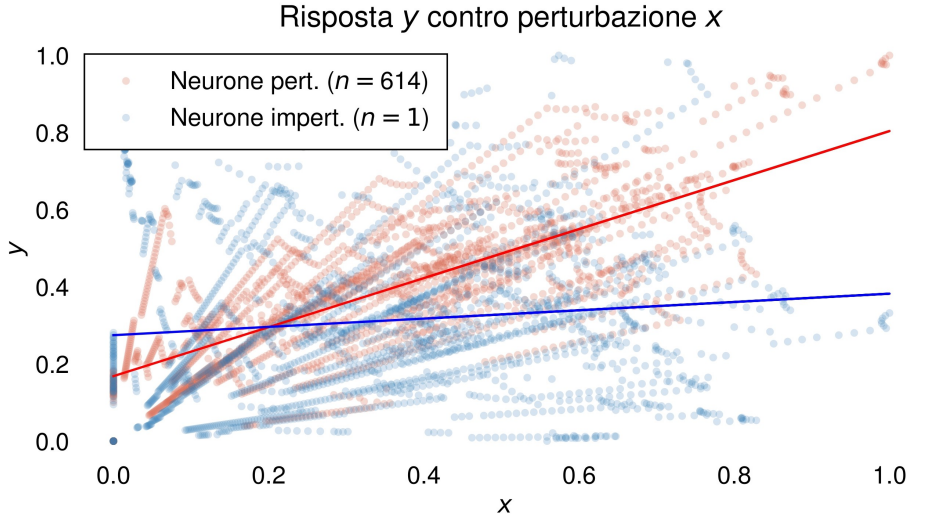


**Figura 3.3:** Test naïf con correzione di Benjamini-Yekutieli.

Cor.	VP	FP	FN	VN
-	20	751 ± 17	0	229 ± 17
H	20	567 ± 23	0	413 ± 23
BY	20	659 ± 18	0	320 ± 18

**Tabella 3.1:** Risultati finali per il test naïf.

Come si era già anticipato nei capitoli precedenti, il tentativo di inferire la presenza (o meno) di un effetto significativo della perturbazione sui singoli neuroni fallisce nella maggior parte dei casi, producendo un numero elevato di falsi positivi. In seguito alla correzione per comparazioni multiple, questo numero si riduce, ma non in maniera così considerevole (-24% per il metodo di Hochberg, -12% per il metodo di Benjamini-Yekutieli), risultando ancora



**Figura 3.4:** Scatter plot di  $\mathbf{y}^{(n)}$  contro  $\mathbf{x}$  per  $n = 1$  (neurone imperturbato) e  $n = 614$  (neurone perturbato).

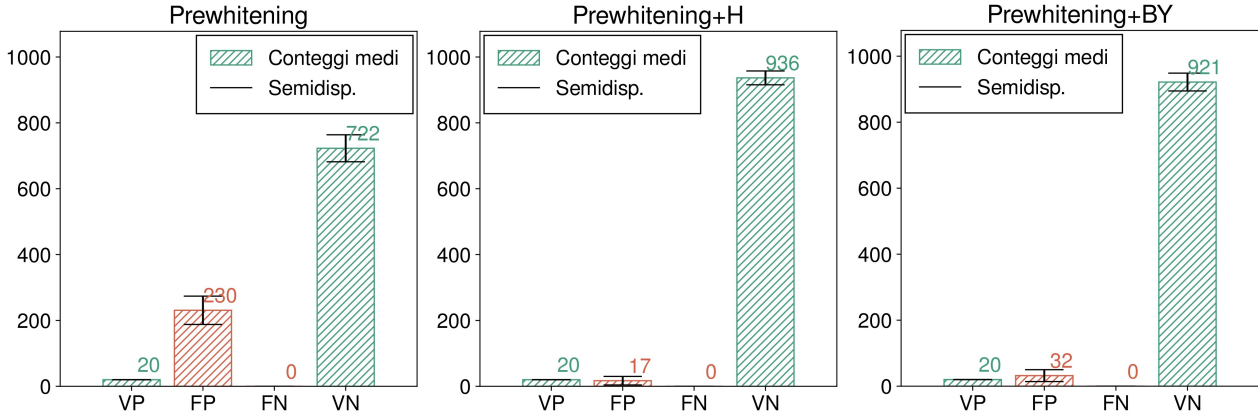
in un conteggio eccessivamente elevato. Lo scatter plot di  $\mathbf{y}^{(n)}$  contro  $\mathbf{x}$  risulta particolarmente articolato e pare essere composto da un fascio di rette centrate indicativamente nell'origine. Euristicamente questo può essere compreso osservando cosa succede in un intervallo di tempo sufficientemente breve in cui si verifichi sia una spike che un impulso di corrente: localmente la convoluzione dei segnali con la risposta del calcio avrà la forma di un esponenziale decrescente con una stessa costante di tempo  $\tau$ :

$$x(t) \sim a + be^{-\frac{t-t_1}{\tau}} \quad y(t) \sim c + de^{-\frac{t-t_2}{\tau}} = c + \frac{d}{b} e^{\frac{t_2-t_1}{\tau}} [x(t) - a] = mx(t) + q$$



Il coefficiente angolare di queste rette è dunque controllato dalla differenza temporale tra gli istanti in cui spike e stimolo si presentano e da questo emerge il fascio osservato.

### 3.1.2 Analisi dei segnali corretti tramite prewhitening



**Figura 3.5:** Prewhitening senza correzioni.

**Figura 3.6:** Prewhitening con correzione di Hochberg.

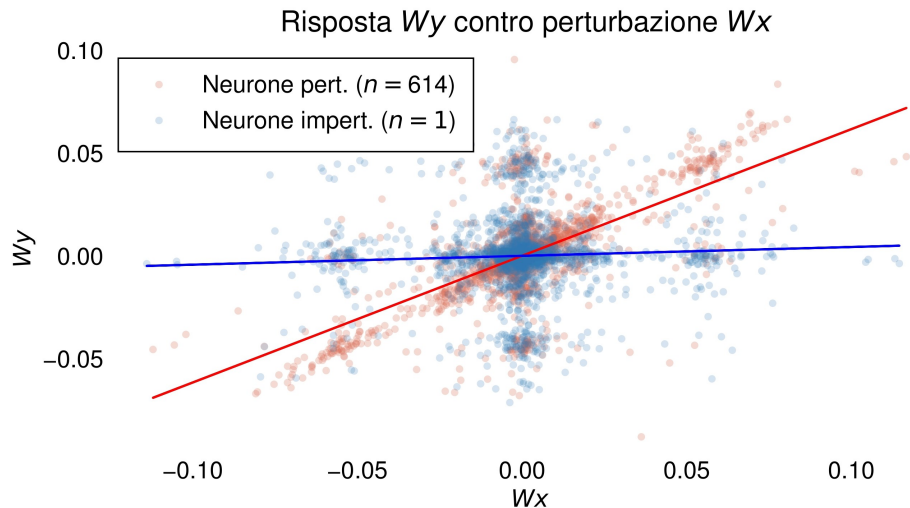
**Figura 3.7:** Prewhitening con correzione di Benjamini-Yekutieli.

Cor.	VP	FP	FN	VN
-	20	231 ± 12	0	723 ± 12
H	20	17 ± 4	0	937 ± 4
BY	20	32 ± 5	0	922 ± 5

**Tabella 3.2:** Risultati finali per il test sui dati autocorrelati.

L'effetto dell'autocorrelazione delle serie temporali sull'inferenza ottenuta è molto importante: rispetto al test naïf senza correzioni per comparazioni multiple il numero di falsi positivi si è ridotto a meno di un terzo. Con le correzioni, poi, gli errori di tipo I calano a meno del 5% rispetto ai corrispettivi valori naïf di Tabella 3.1. Come già si era fatto notare l'autocorrelazione delle serie temporali rappresenta, già a priori, uno dei maggiori ostacoli nell'analisi dei risultati.

Quanto ottenuto mostra quindi come la correzione di tale aspetto possa essere fondamentale per un'indagine statistica che debba operare con serie di dati autocorrelate. Da questa analisi emerge anche l'efficacia delle correzioni per comparazioni multiple, le quali riducono a livelli estremamente bassi il tasso di falsi positivi ottenuti. La correzione di Benjamini-Yekutieli si dimostra leggermente meno conservativa rispetto a quella di Hochberg, ma risulta comunque estremamente efficace. Si fa notare, qui e per i grafici a seguire, che nelle serie in cui il segnale viene corretto tramite prewhitening, al netto del fatto che si stanno guardando dei conteggi medi, generalmente la somma delle entrate non dà 1000. Questo è dovuto al fatto che qualora un neurone non emetta un potenziale d'azione la



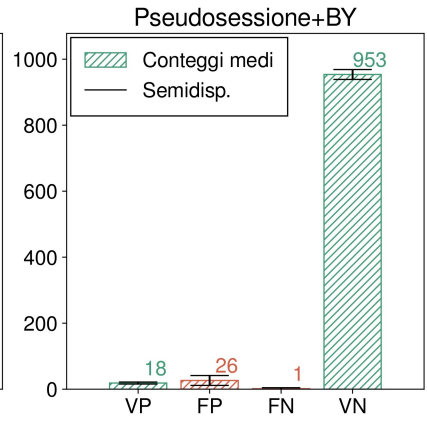
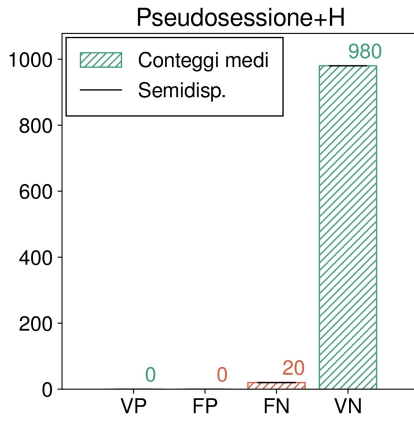
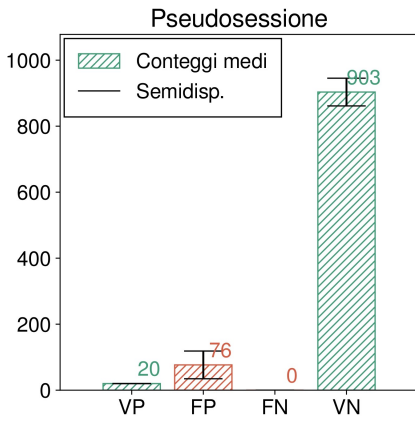
**Figura 3.8:** Scatter plot di  $\mathbf{W}^{(n)}\mathbf{y}^{(n)}$  contro  $\mathbf{W}^{(n)}\mathbf{x}$  per  $n = 1$  (neurone imperturbato) e  $n = 614$  (neurone perturbato).

convoluzione della serie degli spike restituisce ovviamente un vettore nullo (perché già era nulla  $\mathbf{s}$ ). Non è possibile in questi casi, come si intuisce, applicare la correzione di prewhitening. Nondimeno, il numero di neuroni che non emettono la spike è estremamente ridotto rispetto al totale e si concentra generalmente tra i neuroni di inibizione (che come si vede in Figura 2.4 presentano un'attività ridotta rispetto a quelli di eccitazione)

### 3.1.3 Analisi tramite modelli nulli - senza correzione di prewhitening

#### Pseudosessione

Per generare modelli nulli secondo il metodo della pseudosessione, come già spiegato nel Capitolo 2, si devono produrre delle perturbazioni alternative secondo una distribuzione di probabilità nota. In questa simulazione si sono generati  $N_{pse} = 5000$  stimoli fittizi e conseguentemente altrettanti modelli nulli.



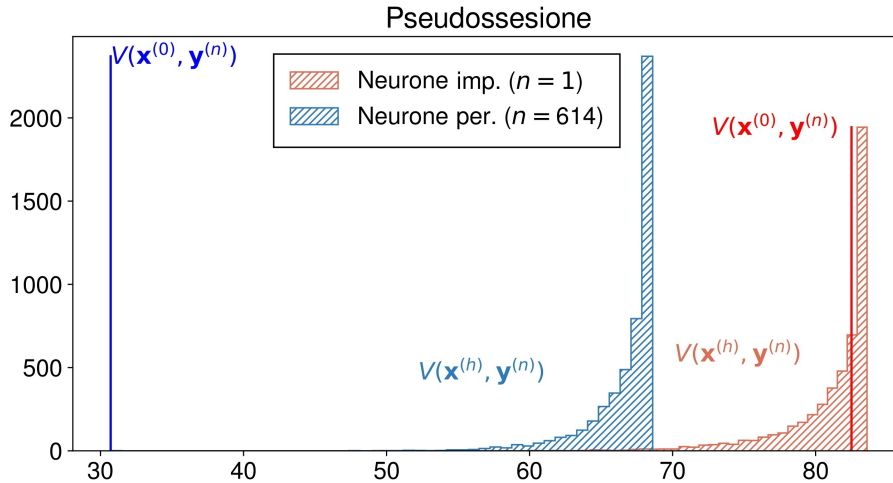
**Figura 3.9:** Modelli nulli: pseudosessione senza correzione.

**Figura 3.10:** Modelli nulli: pseudosessione con correzione di Hochberg.

**Figura 3.11:** Modelli nulli: pseudosessione con correzione di Benjamini-Yekutieli.

Cor.	VP	FP	FN	VN
-	20	76 ± 12	0	893 ± 12
H	0	0	20	980
BY	18 ± 1	26 ± 4	1 ± 1	954 ± 4

**Tabella 3.3:** Risultati finali per il metodo della pseudosessione.



**Figura 3.12:** Distribuzione nulla della statistica di test  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$  per un neurone imperturbato ( $n = 1$ ) ed uno perturbato ( $n = 614$ ) per il metodo della pseudosessione.

### Traslazione Lineare

Con il metodo della traslazione lineare i modelli nulli vengono generati confrontando i primi  $\sigma$  punti di  $\mathbf{y}^{(n)}$  con un blocco di  $\sigma$  punti contigui di  $\mathbf{x}$  traslato ogni volta di uno. In totale, quindi, i modelli nulli che vengono generati, avendo scelto  $\sigma = N'_t/2 = 1500$ , sono per l'appunto  $N'_t - \sigma = \sigma$ .

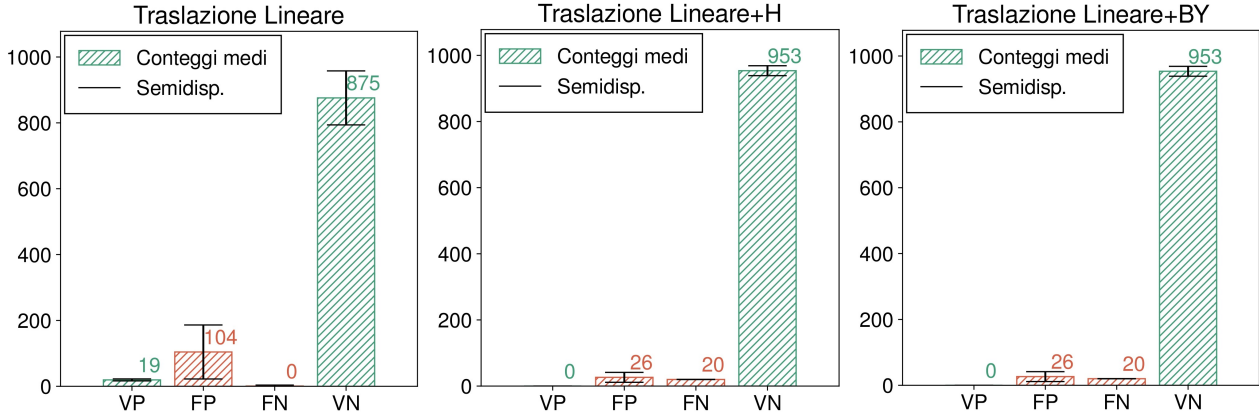


Figura 3.13: Modelli nulli: traslazione lineare senza correzioni.

Figura 3.14: Modelli nulli: traslazione lineare con correzione di Hochberg.

Figura 3.15: Modelli nulli: traslazione lineare con correzione di Benjamini-Yekutieli.

Cor.	VP	FP	FN	VN
-	19 ± 1	104 ± 24	1 ± 1	88 ± 24
H	0	26 ± 4	20	954 ± 4
BY	0	26 ± 4	20	954 ± 4

Tabella 3.4: Risultati finali per il metodo della traslazione lineare.

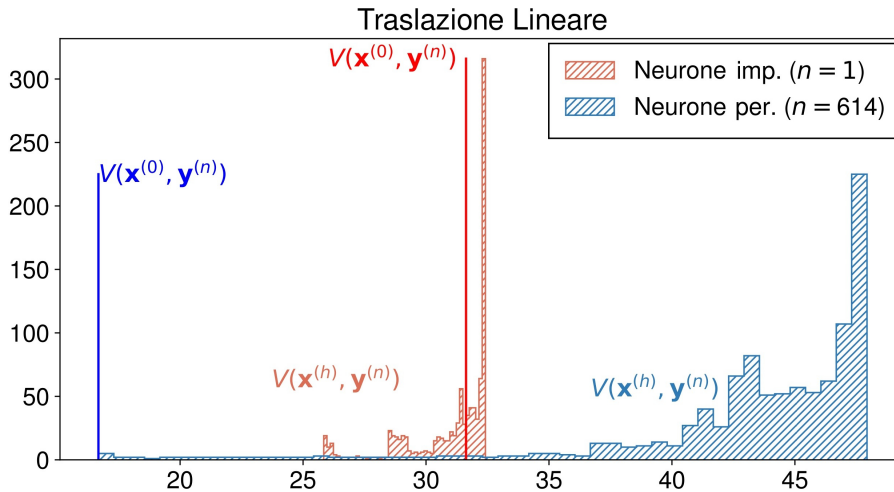
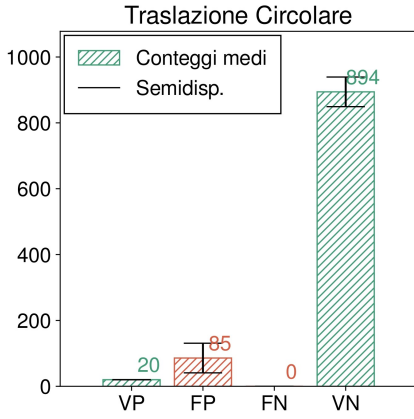


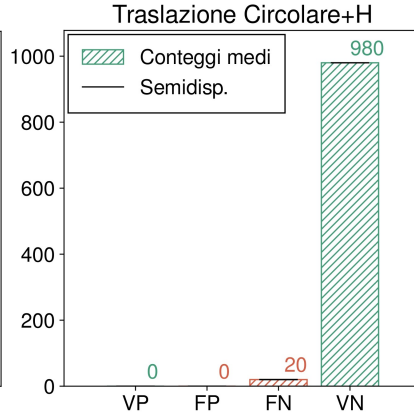
Figura 3.16: Distribuzione nulla della statistica di test  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$  per un neurone imperturbato ( $n = 1$ ) ed uno perturbato ( $n = 614$ ) per il metodo della traslazione lineare.

### Traslazione Circolare

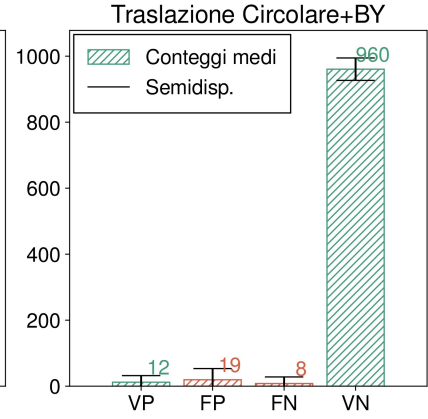
Con il metodo della traslazione circolare i modelli nulli vengono generati confrontando  $\mathbf{y}^{(n)}$  con una versione di  $\mathbf{x}$  ottenuta traslando circolarmente in avanti ogni volta di un punto in più i suoi elementi. In totale quindi i modelli nulli che vengono generati sono pari a  $N'_t - 1$  (perché una traslazione di  $N'_t$  punti è equivalente ad una traslazione di 0 punti).



**Figura 3.17:** Modelli nulli: traslazione circolare senza correzioni.



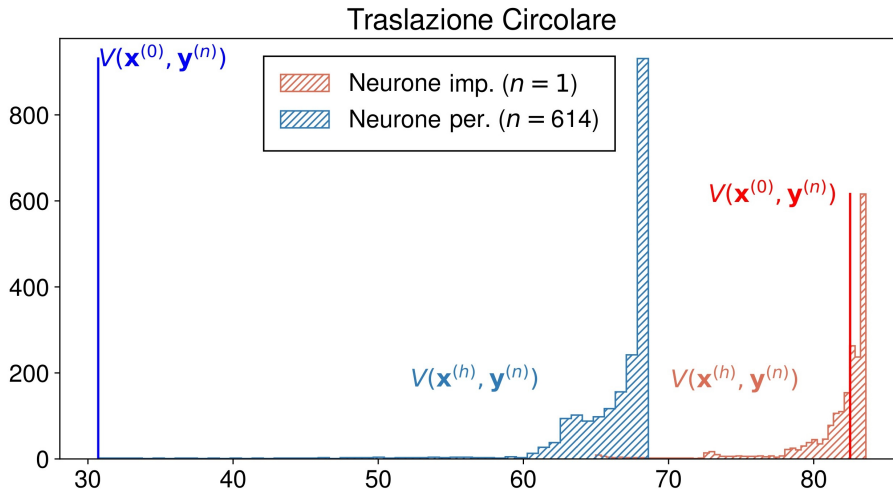
**Figura 3.18:** Modelli nulli: traslazione circolare con correzione di Hochberg.



**Figura 3.19:** Modelli nulli: traslazione circolare con correzione di Benjamini-Yekutieli.

Cor.	VP	FP	FN	VN
-	20	86 ± 13	0	894 ± 13
H	0	0	20	980
BY	12 ± 6	19 ± 10	8 ± 5	961 ± 10

**Tabella 3.5:** Risultati finali per il metodo della traslazione circolare.



**Figura 3.20:** Distribuzione nulla della statistica di test  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$  per un neurone imperturbato ( $n = 1$ ) ed uno perturbato ( $n = 614$ ) per il metodo della traslazione circolare.

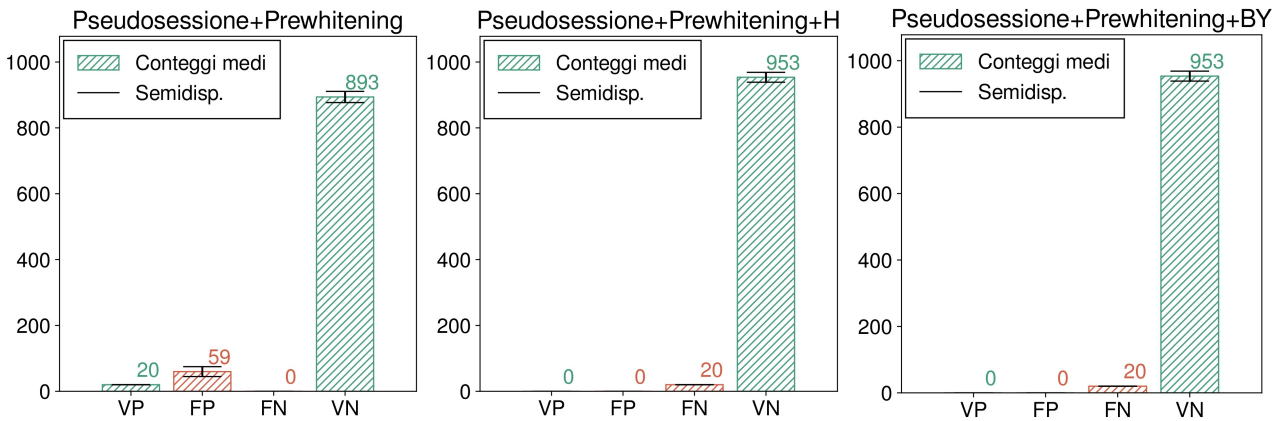
Si è rimandato alla fine di questo paragrafo un commento generale sull'efficacia dei modelli nulli impiegati, giacché emergono diverse caratteristiche comuni. Confrontando il tasso di falsi positivi ottenuto con l'analisi del segnale corretto tramite prewhitening e quello ottenuto impiegando uno qualsiasi dei tre modelli nulli si vede immediatamente come questi ultimi risultino decisamente più efficaci rispetto al primo caso, riducendo almeno della metà il numero di inferenze scorrette. Applicando le correzioni per comparazioni multiple, tuttavia, il discorso pare quasi ribaltarsi: sebbene i falsi positivi siano controllati meglio, ora emerge un numero elevato di falsi negativi rispetto al valore massimo possibile, che è  $N_P = 20$ . Precisamente, nei tre casi la correzione di Hochberg indica come non significativi tutti e 20 i neuroni perturbati, mentre la natura meno conservativa della correzione di Benjamini-Yekutieli fa sì che il numero di falsi negativi sia in generale più basso. In effetti pare che il tasso di falsi negativi cresca al diminuire del numero di modelli nulli impiegato. Questo fatto ha una spiegazione abbastanza semplice: i p-value vengono calcolati come la frazione (sul totale dei modelli nulli generati) di statistiche di test con valore inferiore rispetto a quello della statistica calcolata nella configurazione di partenza. Nel migliore dei casi quindi, il p-value sarà pari a  $1/N_0$ . Giacché i modelli nulli creati sono

dell'ordine delle migliaia e, per la correzione di Hochberg, il fattore moltiplicativo con cui si correggono i p-value è dell'ordine di  $N = N_e + N_i = 1000$  (il numero di ipotesi nulle testate simultaneamente), evidentemente anche quando il p-value non corretto è estremamente significativo la correzione automaticamente ne invalida la significatività. Il problema quindi non risiede tanto nei modelli nulli, di per sé potenzialmente efficaci, quanto piuttosto nel numero di questi ultimi che si è in grado di generare. Si tratta di fatto di un limite computazionale che con le risorse a disposizione non si poteva in alcun modo superare.

### 3.1.4 Analisi tramite modelli nulli - con correzione di prewhitening

Le considerazioni fatte nel precedente paragrafo per ognuno dei tre modelli nulli restano le stesse anche dopo l'introduzione della correzione per prewhitening e pertanto non verranno riproposte. Per un commento generale sull'efficacia dei modelli nulli in questa configurazione si rimanda invece alla fine del presente paragrafo.

#### Pseudosessione



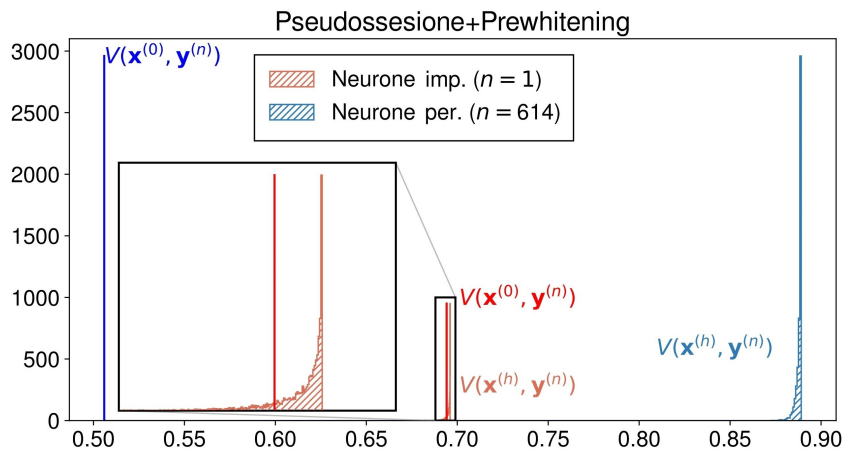
**Figura 3.21:** Modelli nulli: pseudosessione con prewhitening senza correzioni

**Figura 3.22:** Modelli nulli: pseudosessione con prewhitening e correzione di Hochberg

**Figura 3.23:** Modelli nulli: pseudosessione con prewhitening e correzione di Benjamini-Yekutieli

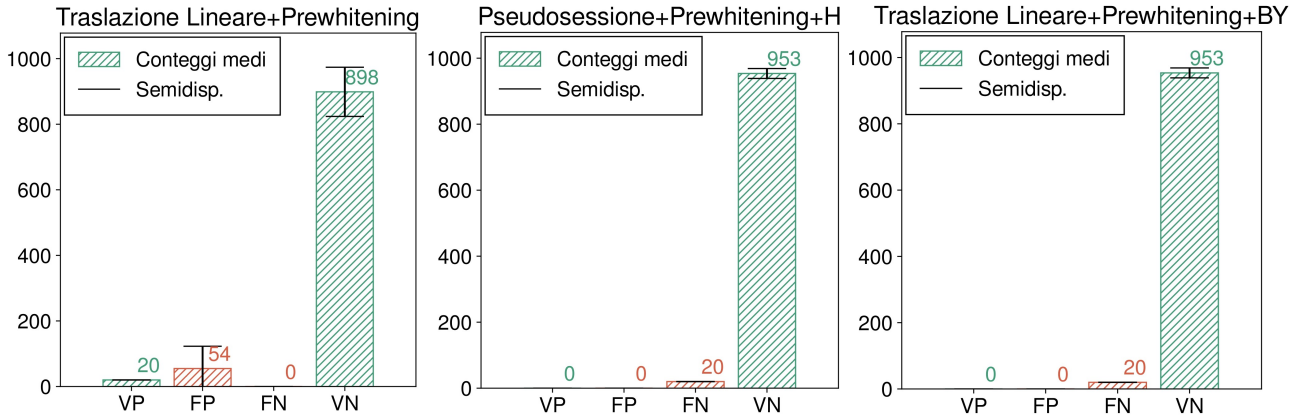
Cor.	VP	FP	FN	VN
-	20	$768 \pm 8$	0	$211 \pm 8$
H	20	$594 \pm 18$	0	$386 \pm 18$
BY	20	$686 \pm 12$	0	$293 \pm 12$

**Tabella 3.6:** Risultati finali per il metodo della traslazione lineare con correzione di prewhitening.



**Figura 3.24:** Distribuzione nulla della statistica di test  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$  per un neurone imperturbato ( $n = 1$ ) ed uno perturbato ( $n = 614$ ) per il metodo della pseudosessione con correzione di prewhitening.

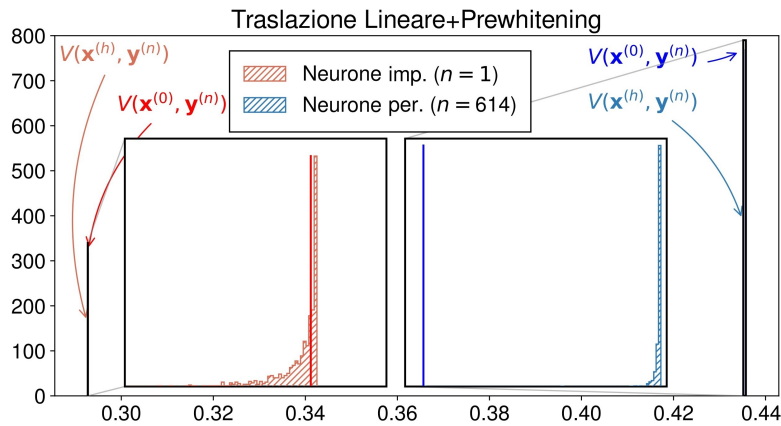
**Traslazione Lineare**



**Figura 3.25:** Modelli nulli: traslazione lineare con prewhitening senza correzioni **Figura 3.26:** Modelli nulli: traslazione lineare con prewhitening e correzione di Hochberg **Figura 3.27:** Modelli nulli: traslazione lineare con prewhitening e correzione di Benjamini-Yekutieli

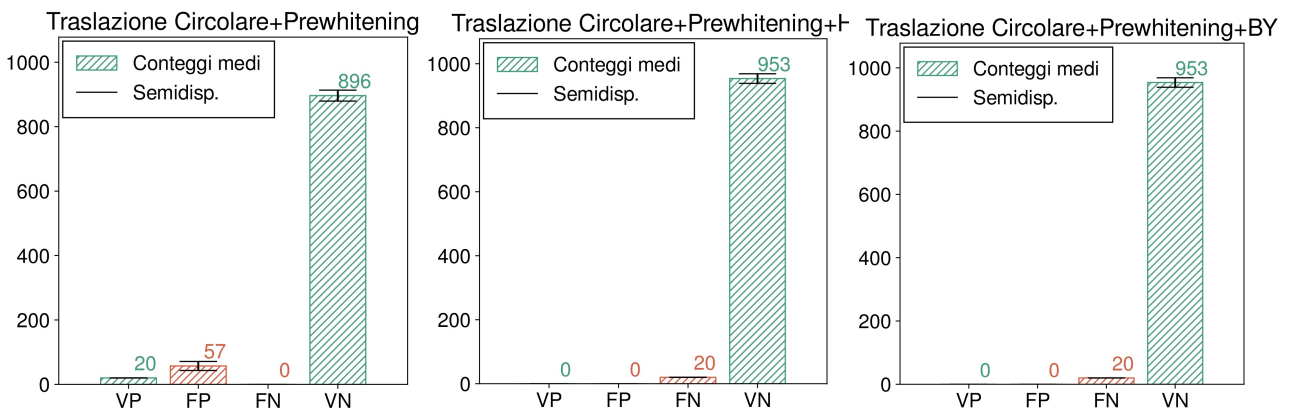
Cor.	VP	FP	FN	VN
-	20	55 ± 20	0	211 ± 22
H	0	0	20	953 ± 4
BY	0	0	20	953 ± 4

**Tabella 3.7:** Risultati finali per il metodo della traslazione lineare con correzione di prewhitening.



**Figura 3.28:** Distribuzioni nulla della statistica di test  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$  per un neurone imperturbato ( $n = 1$ ) ed uno perturbato ( $n = 614$ ) per il metodo della traslazione lineare con correzione di prewhitening.

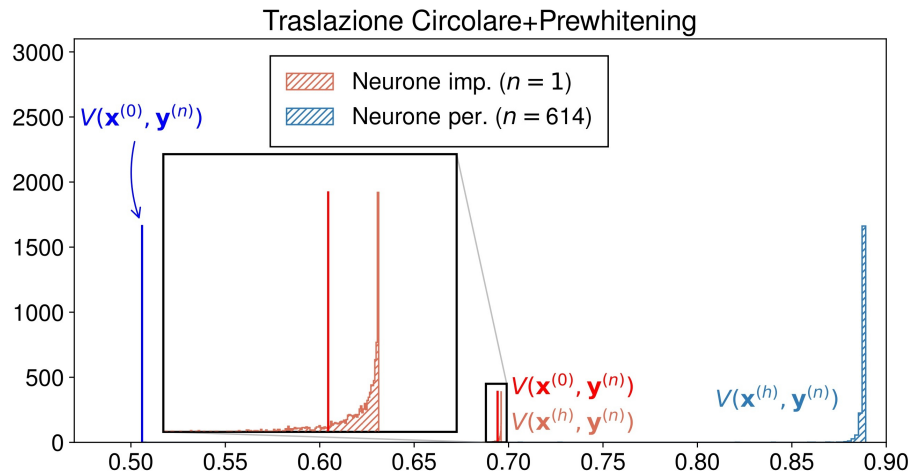
**Traslazione Circolare**



**Figura 3.29:** Modelli nulli: traslazione circolare con prewhitening senza correzioni **Figura 3.30:** Modelli nulli: traslazione circolare con prewhitening e correzione di Hochberg **Figura 3.31:** Modelli nulli: traslazione circolare con prewhitening e correzione di Benjamini-Yekutieli

Cor.	VP	FP	FN	VN
-	20	$57 \pm 4$	0	$896 \pm 5$
H	0	0	20	$954 \pm 4$
BY	0	0	20	$954 \pm 4$

**Tabella 3.8:** Risultati finali per il metodo della traslazione circolare con correzione di prewhitening.



**Figura 3.32:** Distribuzioni nulla della statistica di test  $V(\mathbf{x}^{(h)}, \mathbf{y}^{(n)})$  per un neurone imperturbato ( $n = 1$ ) ed uno perturbato ( $n = 614$ ) per il metodo della traslazione circolare con correzione di prewhitening.

Si ripetono le considerazioni già fatte in precedenza: le correzioni per comparazioni multiple, a causa del numero non sufficientemente elevato di modelli nulli creati, portano ad ottenere un numero elevato di falsi negativi. In questo caso non emerge nemmeno il carattere meno conservativo della correzione di Benjamini-Yekutieli: in qualsiasi caso, per i tre modelli nulli e le due correzioni utilizzate tutti i neuroni perturbati vengono dichiarati come non perturbati. Il problema, di nuovo, non è imputabile all'efficacia dei modelli nulli come metodo in sé, giacché, se si guardano solamente i risultati ottenuti senza le correzioni, si nota che il tasso di falsi positivi è il più basso di tutte le configurazioni studiate. Paragonando poi questi risultati con quelli ottenuti tramite modelli nulli ma senza la correzione di prewhitening si può constatare, una volta di più, come l'autocorrelazione delle serie temporale giochi un ruolo molto importante nell'analisi statistica di questo genere di esperimenti e vada quindi rimossa.





# Capitolo 4

## Conclusioni

In questo lavoro abbiamo indagato l'accuratezza di strumenti statistici basati sulla regressione lineare per identificare i neuroni che rispondono a perturbazioni esterne (stimoli) a partire dai segnali di fluorescenza rilevati nel calcium imaging. Per sondare l'efficacia dei metodi statistici in esame ci siamo basati su simulazioni numeriche della dinamica di una rete neurale in presenza di perturbazioni.

I risultati ottenuti con le nostre simulazioni hanno fatto emergere alcuni aspetti rilevanti nell'analisi statistica di questa tipologia di problemi. In primo luogo, l'evidenza più scontata è che applicare i consueti test statistici direttamente sui dati osservati è estremamente inefficace: il risultato che si ottiene è che una cospicua frazione delle unità, molto maggiore di quella reale, viene dichiarata come perturbata. La causa di ciò è da ritrovare in buona parte nel fatto che le serie di dati che si acquisiscono, sperimentalmente o tramite una simulazione come in questo caso (e come d'altro canto spesso capita quando si ha a che fare con delle serie temporali) sono costituite da punti non indipendenti tra di loro. Questo aspetto allontana il problema dal campo di applicabilità dei test di ipotesi più semplici, che, come si è detto, falliscono. La rimozione dell'autocorrelazione dei dati è sufficiente per ridurre drasticamente il tasso di errori di I specie. Ciò non è tuttavia sufficiente, giacché quando si confronta simultaneamente più di un'ipotesi è necessario adottare le opportune correzioni, al fine di non produrre un numero comunque troppo elevato di questi errori. Una volta applicate le correzioni per comparazioni multiple si è infatti osservato che per i dati ripuliti dall'autocorrelazione (ma questo è vero in una certa misura anche per il caso delle serie ancora autocorrelate) il numero di falsi positivi si attesta a valori effettivamente molto bassi. Per operare un'analisi più sofisticata, si sono infine introdotti i modelli nulli. In effetti la potenza inferenziale di questi metodi è apparsa evidente confrontando il tasso di falsi positivi ottenuti tramite queste tecniche con quelli ottenuti nei test più classici (metodo naïf e correzione tramite prewhitening). È emerso anche come la correzione dell'autocorrelazione prima dell'applicazione dei modelli nulli restituisca risultati ancora più accurati (a riprova di come questo sia un aspetto molto importante del problema). Quando si sono applicate le correzioni per comparazioni multiple si è osservato, tuttavia, l'insorgere di un numero non trascurabile di falsi negativi. Si è ad ogni modo avuta cura di precisare come questi errori emergano non tanto per un limite dei modelli nulli come strumento inferenziale in sé, quanto piuttosto per un vincolo di natura puramente computazionale: i modelli nulli sono troppo ridotti in numero e sarebbe necessario aumentare la lunghezza delle simulazioni per ottenere ensemble nulli più cospicui. Tutto ciò ha ovviamente un costo in termini di elaborazione numerica; qualora esso non possa essere superato, si palesa la necessità di introdurre metodi di correzione ancora più sofisticati.



# Bibliografia

- [1] Anthony N. Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological Cybernetics*, 95:1–19, 2006.
- [2] Shiyi Chen, Zhe Feng, and Xiaolian Yi. A general introduction to adjustment for multiple comparisons. *Journal of thoracic disease*, 9 6:1725–1729, 2017.
- [3] Kenneth D. Harris. Nonsense correlations in neuroscience. *bioRxiv*, 2021.
- [4] E.M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572, 2003.
- [5] Eugene M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, 07 2006.
- [6] Russell A. Poldrack, Jeanette A. Mumford, and Thomas E. Nichols. *Handbook of Functional MRI Data Analysis*. Cambridge University Press, 2011.
- [7] Vinita Vasudevan and M. Ramakrishna. A hierarchical singular value decomposition algorithm for low rank matrices. *ArXiv*, abs/1710.02812, 2017.
- [8] František Váša and Bratislav Mišić. Null models in network neuroscience. *Nature reviews. Neuroscience*, 23(8):493—504, August 2022.
- [9] G. Udny Yule. Why do we sometimes get nonsense-correlations between time-series?—a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1):1–63, 1926.