

UNIVERSITÀ DEGLI STUDI DI FIRENZE
Dipartimento di Statistica “G. Parenti”

Tesi di Dottorato di Ricerca
in
Statistica Applicata

**La rivelazione di segnali transienti tramite
la rappresentazione wavelet discreta**

Candidato: Dott. Leonardo Fabbroni

Coordinatore: Prof. Giovanni M. Marchetti

Relatore: Prof.ssa Marina Vannucci

Correlatore: Prof. Fabio Corradi

Anno Accademico 1999-2000

Indice

1	Introduzione	1
1.1	Un interessante problema	1
1.2	La scelta delle wavelets	3
1.3	La struttura del lavoro	6
2	Metodologie per la rivelazione di eventi transienti	7
2.1	I test proposti in letteratura	9
2.1.1	<i>Bin Counting</i>	10
2.1.2	<i>Norm Filter</i>	10
2.1.3	<i>Norm of the Autocorrelation</i>	11
2.1.4	<i>Slope Filter</i>	11
2.1.5	<i>Offset Filter</i>	12
2.1.6	<i>Peak Correlator</i>	12
2.2	I test wavelet	13
2.2.1	Distribuzioni di probabilità delle statistiche test	23
2.2.2	Test in intervalli non disgiunti	27
2.2.3	Considerazioni pratiche	31
2.3	Test in presenza di rumore correlato	36
3	Rivelazione di segnali gravitazionali da supernovae	41
3.1	Criteri di valutazione delle prestazioni	41
3.1.1	Test del rapporto di verosimiglianza	44
3.1.2	Distanze soglia	47
3.2	Le prestazioni dei test wavelet	48
3.2.1	Dipendenza dall'ampiezza α del test	65
3.2.2	Dipendenza dalla larghezza n dell'intervallo	66

3.2.3	Potenza dei test wavelet	70
3.3	Prestazioni nel caso di rumore correlato	74
3.4	Densità di probabilità di rivelazione	83
4	Conclusioni	89
4.1	Discussione	89
4.2	Problemi aperti e sviluppi futuri	92
A	Le Wavelets	95
A.1	Trasformate Wavelet	95
A.1.1	La trasformata wavelet discreta	98
	Bibliografia	105

Capitolo 1

Introduzione

La statistica è una scienza multidisciplinare che costruisce le proprie fondamenta col rigore del linguaggio matematico per trovare applicazione in moltissimi campi assai diversi l'uno dall'altro, dalle scienze naturali a quelle sociali.

La fisica sperimentale rientra in questa vasta schiera di domini nei quali la statistica trova un utilizzo abituale. Questo legame è enfatizzato qualora si ricordi che gli albori delle scienze statistiche sono legati ai nomi di Newton, Bernoulli, Gauss, scienziati che hanno fatto al tempo stesso la storia della fisica.

L'analisi dei dati in un esperimento di fisica necessita sempre di opportune metodologie statistiche per trattare l'aleatorietà che caratterizza qualsiasi tipo di misura. Fondamentale è poi l'apporto della statistica nel confronto continuo tra modello ed esperienza che costituisce l'essenza del processo deduttivo-induttivo alla base di tutte le scienze sperimentali, fisica inclusa.

Talvolta capita che le condizioni sperimentali siano tali da richiedere un'innovazione o un riadattamento peculiare delle metodologie statistiche esistenti. Ci è sembrato che questo potesse essere il caso per l'esperimento VIRGO in allestimento nella piana di Cascina (Pi) ed in procinto di entrare in funzione per una prima fase di caratterizzazione strumentale nel corso del 2001 .

1.1 Un interessante problema

VIRGO è uno strumento interferometrico di cospicue dimensioni progettato per la rivelazione di onde gravitazionali, un fenomeno fisico previsto dalla teoria della relatività generale, ma non ancora direttamente osservato. L'analisi

dei dati da effettuare riguarda sostanzialmente la serie storica delle variazioni di distanza tra due masse di riferimento, come descritto in [45]. Benché lo strumento sia in fase avanzata di realizzazione, la definizione di metodologie per il trattamento dei dati che da esso verranno prodotti è tuttora in pieno fermento e alla ricerca di nuovi contributi.

A rendere interessante il problema da un punto di vista statistico contribuiscono una serie di fattori:

- una complessa componente di rumore che rende particolarmente ardua la rivelazione dei segnali gravitazionali. Al rumore strumentale contribuisce una vasta e diversificata varietà di sorgenti tra cui, per citarne alcune: il tremolio microsismico locale, la statistica poissoniana dei fotoni dei laser impiegati nella misura¹, l'agitazione termica delle varie parti dello strumento, le variazioni di disposizione delle masse circostanti². Per alcune delle previste sorgenti di rumore vengono predisposte delle misure periodiche da registrare in serie storiche covariate, per altre ciò non è possibile. Nel complesso ci si aspetta un rumore fortemente correlato, presumibilmente stazionario solo per intervalli limitati di tempo e per il quale non è detto che sia del tutto accettabile l'ipotesi di gaussianità;
- segnali da rivelare che in molti casi sono conosciuti solo approssimativamente e per i quali conseguentemente non sono applicabili metodologie ottimali (come il test del rapporto di verosimiglianza) o strettamente parametriche;
- un'enorme mole di dati acquisiti a 10-20 KHz, cioè uno ogni 0.1-0.05 mililisecondi, per diversi mesi consecutivi; tali dati devono essere analizzati in tempi confrontabili con quelli di acquisizione, con conseguente richiesta di metodologie computazionalmente non intensive, a volte a scapito di altre loro desiderabili proprietà.

Di fatto va poi sottolineato che gli eventi selezionati in prima battuta come possibili fenomeni di origine gravitazionale dovranno essere successivamente

¹L'energia della luce è trasportata in pacchetti discreti detti *fotoni*. L'intensità della luce misurata da un rivelatore è legata al numero di fotoni che nell'unità di tempo arrivano sul rivelatore stesso. Tale numero è in realtà una variabile casuale con distribuzione di Poisson.

²Con la conseguenza di variare l'attrazione gravitazionale esercitata sulle masse di riferimento.

validati per mezzo di un'analisi di coincidenze temporali con serie indipendenti di osservazioni, provenienti da altri strumenti per la rivelazione di onde gravitazionali ed osservazioni astronomiche. Anche questa fase rientra a pieno titolo nelle competenze della statistica e può presentare problemi assai complessi e aperti a innovazioni metodologiche, come, ad esempio, l'analisi congiunta e contemporanea di più serie storiche che misurano lo stesso fenomeno utilizzando diverse strumentazioni o la meta-analisi di databases di eventi molto diversi ma potenzialmente correlati.

1.2 La scelta delle wavelets

Molti sono i tipi di approccio che si possono scegliere per l'analisi di serie storiche a seconda del contesto e delle finalità dell'analisi che si intende svolgere. L'interesse può essere focalizzato sulla capacità di prevedere l'andamento futuro della serie, come spesso accade in campo economico. Nelle discipline fisiche invece si è di norma interessati ad interpretare l'andamento passato della serie e scomporlo in componenti con distinti e specifici significati nel contesto in cui si opera. Anche nella scelta dei modelli da adottare si fronteggiano filosofie diverse, da modelli guidati prevalentemente dai dati (ARMA, GARCH etc., si veda ad esempio [27]) a modelli costruiti sulla base di qualche schematizzazione a priori del fenomeno oggetto di studio, fino ad arrivare ad un'inclusione delle conoscenze a priori in qualche forma di modello bayesiano (come proposto ad esempio nel caso dei Dynamic Models, [49]). Naturalmente la distinzione tra i vari possibili approcci non è marcata da confini invalicabili, a volte si può trattare di sfumature piuttosto labili e spesso è conveniente condurre in parallelo analisi con metodologie e assunzioni diverse.

Nel campo delle scienze fisiche modelli che consentano una scomposizione della serie storica in termini di un insieme di elementi di chiara interpretazione sono quasi sempre la scelta da preferire, visto che le finalità sono assai più spesso interpretative che di previsione.

Inoltre in questo campo per consuetudine c'è una diffusa tendenza a privilegiare l'analisi nel dominio delle frequenze, ricorrendo alle densità spettrali, piuttosto che non direttamente nel dominio temporale. In effetti questa propensione manifesta l'abitudine a trattare serie storiche stazionarie del secondo

ordine, quelle per cui ha senso parlare di densità di potenza spettrale.

D'altra parte in presenza di segnali impulsivi, specie se di breve durata, o comunque di fenomeni che non si manifestino con una certa regolarità, il dominio temporale è di solito più adeguato.

I due dominî sono perfettamente complementari: rappresentando una serie storica nel suo dominio temporale non si evidenziano i contributi dati dalle varie frequenze, mentre nella sua rappresentazione di Fourier si perde ogni possibile localizzazione temporale. Nasce così l'esigenza di ricorrere ad altre trasformazioni che siano in grado di fornire informazioni su eventuali periodicità senza perdere la capacità di seguire nel tempo i possibili cambiamenti delle caratteristiche della serie analizzata. La trasformata di Gabor (o, più in generale, le Short Time Fourier Transform), la trasformata wavelet continua, la distribuzione di Wigner-Ville e molte altre sono tutti strumenti matematici che offrono la possibilità di rispondere a questa esigenza di rappresentare la serie storica oggetto di studio simultaneamente nel dominio dei tempi e in quello delle frequenze; si veda [22] o [41].

Alla luce di queste considerazioni l'utilizzo delle rappresentazioni appena citate risulta particolarmente promettente per l'analisi di dati come quelli prodotti nell'esperimento VIRGO. Infatti ad una varietà di caratteristiche da cogliere a frequenze diverse si affianca la necessità di una localizzazione temporale, come conseguenza di contributi (di rumore e di segnale) di durata limitata e di probabili evoluzioni temporali della componente di rumore localmente considerata come stazionaria.

Molte sono le caratteristiche e le proprietà che differenziano tra loro le varie rappresentazioni tempo-frequenza ed elencarne vantaggi e svantaggi va al di là degli scopi di questa introduzione. Tuttavia deve essere ricordato che ogni rappresentazione di questo tipo è sempre soggetta al principio di indeterminazione:

$$\Delta t \cdot \Delta f \geq \text{cost}$$

ossia non può arbitrariamente migliorare la propria risoluzione temporale Δt se non a scapito della propria risoluzione in frequenza Δf (e viceversa). Le diverse rappresentazioni citate si distinguono anche per le differenti modalità di risoluzione con cui ottemperano al principio di indeterminazione.

Le trasformate wavelet offrono la proprietà di incrementare la risoluzione in

frequenza alle basse frequenze e di migliorare viceversa la risoluzione temporale alle alte frequenze (tipiche di fenomeni rapidi). In effetti questa proprietà sembra adattarsi perfettamente al problema della ricerca di segnali (o di rumori) impulsivi di durata limitata in un sottofondo di rumore lentamente variabile nel tempo, come in VIRGO ci si aspetta che accada, ad esempio, per i segnali originati in esplosioni di supernovae.

Altro elemento a favore delle wavelets è il loro consolidato impiego in ambito statistico non parametrico: si consulti [3] o [44]. In questo contesto le wavelets vengono sfruttate per imporre ben determinate caratteristiche di regolarità a distribuzioni o funzioni da adattare a dati empirici: ad esempio si sfrutta questa metodologia per la stima non parametrica di densità di probabilità o di curve di regressione, [33], [2]. Nel caso di VIRGO questo tipo di impiego è pertinente in tutte quelle (molte) situazioni in cui non sia disponibile un'affidabile modellistica parametrica di ciò che si sta cercando, sia esso un segnale gravitazionale o una caratteristica del rumore sperimentale.

Inoltre l'uso di metodi non parametrici basati sulle wavelets è stato proposto anche per un campo classico come quello della stima di densità di potenza spettrale per serie storiche stazionarie, [25], a fianco di una grande varietà di altri metodi ormai ben consolidati. Ancora più interessanti sono i recenti tentativi di utilizzare le wavelets per la costruzione di una sorta di spettro evolutivo per serie storiche caratterizzate da qualche forma di locale stazionarietà, [36], [20], come sarebbe lecito attendersi dal rumore lentamente variabile di VIRGO.

Infine un altro vantaggio offerto dalle trasformate wavelet è quello di essere calcolabili per mezzo di algoritmi computazionalmente efficienti, [34], e quindi di mantenere in termini accettabili i tempi di calcolo anche per cospicui flussi di dati. Va aggiunto che alcune varianti, quali la trasformata wavelet discreta o la decomposizione in pacchetti wavelet, offrono delle rappresentazioni della serie analizzata convenientemente "parsimoniose": in questi casi la serie viene scomposta su una base ortonormale completa, a differenza del caso della trasformata wavelet continua o di altre rappresentazioni, come la Wigner-Ville, in cui la scomposizione avviene su una base sovracompleta (cioè con una certa ridondanza). Benchè le rappresentazioni su basi sovracomplete siano tipicamente di maggiore intelligibilità a fini di interpretazione, la parsimonia di

coefficienti offerta dalle basi complete diviene una proprietà particolarmente attraente quando ci si trovi a fronteggiare delle moli di dati enormi come quelle prodotte da VIRGO.

1.3 La struttura del lavoro

Nel capitolo 2 viene definito con rigore formale il tipo di problema che si intende affrontare in questo lavoro. Alcune delle soluzioni proposte in letteratura per questo problema sono discusse nella sezione 2.1. Nella successiva sezione 2.2 vengono introdotte alcune originali metodologie per la soluzione del problema posto, costruite grazie allo strumento matematico della rappresentazione wavelet discreta. Le proprietà teoriche ed alcuni aspetti pratici di queste metodologie originali vengono approfonditi nelle sezioni 2.2.1, 2.2.2 e 2.2.3. Nella sezione 2.3 si estende il campo di applicazione delle metodologie proposte dal semplice caso di rumore bianco al caso, assai più generale, di rumore correlato e localmente stazionario.

Nel capitolo 3 si applicano i risultati del capitolo 2 ad uno specifico problema applicativo, quello della rivelazione di una particolare classe di segnali gravitazionali grazie all'esperimento VIRGO. Nella sezione 3.1 si introducono alcuni criteri di valutazione delle metodologie utilizzate nel contesto della rivelazione dei segnali gravitazionali. I risultati ottenuti per questa specifica applicazione con le metodologie proposte in questo lavoro vengono discussi nella sezione 3.2, confrontandoli con i risultati ottenibili con le metodologie presentate in letteratura. Nelle sezioni 3.2.1 e 3.2.2 si valuta l'influenza delle scelte operate per alcuni parametri, mentre nella sezione 3.3 si affronta il medesimo problema applicativo tenendo però conto della struttura di correlazione del rumore strumentale. La sezione 3.4 pone il problema di combinare le prestazioni ottenute dalle varie metodologie con le esigenze effettive del problema applicativo oggetto di studio.

Infine nella sezione 4.1 si discutono i risultati ottenuti nel presente lavoro, sia dal punto di vista fisico che da quello statistico. La sezione 4.2 ricorda alcuni problemi incontrati che necessitano di ulteriori approfondimenti e indica alcune linee di ricerca da seguire per futuri lavori. L'appendice A conclude la tesi, fornendo una semplice introduzione alla teoria delle wavelets.

Capitolo 2

Metodologie per la rivelazione di eventi transienti

Il problema che ci si propone di affrontare in questo capitolo è quello di rivelare, su base statistica, la possibile sovrapposizione di un segnale di breve durata s_k in una serie storica di rumore X_k gaussiano e localmente stazionario in senso debole per tempi più lunghi della durata dell'eventuale segnale. Si assume che la conoscenza sul segnale da ricercare sia molto ridotta e che per tale motivo si sia costretti a rinunciare all'implementazione di un test ottimale come quello effettuato sulla base del rapporto di verosimiglianza.

L'analisi riguarda dunque una serie storica a tempi discreti Y_k ($k \in \mathbb{Z}$) ottenuta dal campionamento di una serie continua ad intervalli di tempo t_c . Per i dati della serie si adotta il seguente elementare modello:

$$Y_k = \begin{cases} X_k, & \text{se il segnale non è presente (ipotesi nulla } H_0); \\ X_k + s_k, & \text{altrimenti (ipotesi alternativa } H_1). \end{cases} \quad (2.1)$$

Si suppone di conoscere approssimativamente la durata del segnale s e la sua larghezza di banda, limitata ad una frequenza di taglio superiore sensibilmente più piccola della frequenza di campionamento $f_c = 1/t_c$. A parte queste assai scarse informazioni non si dispone di nessuna conoscenza sull'andamento temporale del segnale. Il rumore X , chiaramente di tipo additivo nel modello proposto, è supposto essere gaussiano e localmente stazionario, termine col quale si intende che la corrispondente matrice di varianza e covarianza può essere soggetta ad una lenta evoluzione temporale. Inizialmente sarà considerato semplicemente un rumore bianco con varianza σ_X^2 incognita; in una seconda fase si affronterà il caso di un rumore correlato.

Questo tipo di problema risulta di rilevante interesse nel contesto dell'esperimento VIRGO. Infatti allo stato attuale l'astrofisica teorica non è in grado di proporre modelli attendibili ed accurati per un importante tipo di segnale che ci si propone di rivelare con VIRGO: le onde gravitazionali originate nel corso dell'esplosione con cui stelle di grande massa (maggiore di 7-8 masse solari) pongono fine alla propria esistenza. Si tratta di un evento impulsivo con una durata limitata approssimativamente a qualche decina di millisecondi. Per contro ci si aspetta che il rumore dello strumento, benché non stazionario, abbia una componente prevalente con caratteristiche variabili su tempi che vanno da qualche minuto a molte ore; in questa situazione, localmente, ha senso parlare di densità di potenza spettrale per il processo stocastico del rumore strumentale.

Per la verità esiste anche una serie di alterazioni aleatorie della configurazione strumentale che originano una componente impulsiva e sporadica del rumore. Si pensi ad esempio al caso dei microscopici cedimenti meccanici di alcune parti delle sospensioni che sostengono il peso delle masse di riferimento, per isolarle dalle vibrazioni del terreno su cui è installato lo strumento. Sebbene queste fenomenologie rientrino senza dubbio nella categoria del rumore, si preferisce indicarle piuttosto col termine di "voci strumentali" per distinguerle dalla componente localmente stazionaria del rumore. In effetti per quanto riguarda gli scopi di questa tesi le voci strumentali vengono considerate come veri e propri segnali da rivelare. Infatti nel caso in cui anche per esse la modellistica fisica sia insoddisfacente, la loro individuazione è perseguibile con le stesse metodologie proponibili per i segnali gravitazionali transienti indicati in precedenza.

Per quanto appena detto dunque il problema che ci si pone è esteso alla ricerca della possibile presenza di un evento transiente di limitata durata senza porre vincoli sulla sua natura, sia esso frutto del passaggio di una vera onda gravitazionale o la manifestazione di una voce strumentale. Si rimanda ad una eventuale fase successiva il tentativo di distinguere tra una di queste due possibilità e quella di una peculiare realizzazione della componente quasi stazionaria del rumore, indicata con la significativa locuzione "falso allarme".

Dato il modello 2.1 proposto per la serie storica da analizzare e vista l'assen-

za di conoscenza a priori sulla forma del possibile segnale s , è nella categoria dei test di ipotesi non parametrici o semi-parametrici che debbono essere ricercate le metodologie atte a discriminare tra l'ipotesi nulla H_0 di solo rumore quasi stazionario nei dati e quella alternativa H_1 in cui un transiente si sovrappone al rumore stesso. Naturalmente è auspicabile trovare tests che garantiscano un'elevata potenza per il maggior numero possibile di segnali. In altre parole un buon test per questo tipo di problema dovrebbe essere generalmente robusto rispetto all'andamento temporale di ogni possibile transiente; è inutile garantirsi elevate potenze solo per alcuni particolari transienti a scapito della potenza per altri segnali, dal momento che le conoscenze a priori sono troppo ridotte per esprimere delle preferenze.

2.1 I test proposti in letteratura

Le soluzioni proposte per il problema descritto in precedenza si basano su statistiche test calcolate per una “finestra” di dati contigui nella serie storica. Il valore ottenuto da tali statistiche viene quindi confrontato con un corrispondente valore di soglia individuato in funzione del massimo numero di falsi allarmi, generati dalla componente stazionaria del rumore, che mediamente si è disposti a tollerare nell'unità di tempo. Questa procedura viene quindi ripetuta su una sequenza di finestre con istante iniziale crescente, al fine di coprire i dati dell'intera serie storica.

Si deve poi aggiungere che, in generale, nei lavori che hanno trattato l'argomento della rivelazione di segnali transienti nel contesto dell'esperimento VIRGO, è sempre stata fatta l'assunzione di rumore bianco, cioè non correlato, e con varianza nota. È chiaro che questa assunzione non è altro che una comoda semplificazione, in netto contrasto con la situazione reale in cui ci si attende che il rumore di VIRGO presenti una struttura di correlazione assai complessa. Questa semplificazione tuttavia risulta più accettabile considerando che si prevede di sottoporre i dati originali di VIRGO ad opportune procedure di *whitening* che trasformino la serie storica in modo da eliminare la correlazione nella componente di rumore; esempi di procedure di questo tipo, fondate su un modello autoregressivo eventualmente adattivo per il rumore strumentale, sono descritte in [13] o in [1].

A fronte di queste caratteristiche comuni, le soluzioni proposte si differenziano per la statistica utilizzata nel test di ipotesi da effettuare sui dati di una singola finestra. Nel seguito vengono descritte alcune possibili scelte per la statistica da utilizzare; si assume che la finestra comprenda n dati consecutivi della serie storica oggetto di analisi e che, sotto l'ipotesi nulla, tali dati siano la realizzazione di variabili X_k , $k \in \{1, 2, \dots, n\}$, indipendenti ed identicamente distribuite secondo una distribuzione normale di media nulla e varianza σ_X^2 .

2.1.1 *Bin Counting.*

Per la classe di test descritta in precedenza, la statistica probabilmente più intuitiva è quella che cerca di contare quanti sono, tra gli n dati sotto esame, quelli che assumono valori distanti dalla media 0 più di r volte il valore della standard deviation σ_X . Questo approccio viene in effetti proposto in [4]. Si introducono le variabili aleatorie:

$$Z_k = \begin{cases} 1, & \text{se } |Y_k| \geq r\sigma_X; \\ 0, & \text{altrimenti.} \end{cases} \quad (2.2)$$

Sotto ipotesi nulla anche queste variabili come le Y_k sono indipendenti e identicamente distribuite, ma con distribuzione di Bernoulli con probabilità:

$$p_r = \frac{2}{\sqrt{2\pi}} \int_r^{+\infty} e^{-\frac{x^2}{2}} dx, \quad (2.3)$$

facilmente calcolabile per ogni fissato valore di r a partire dalla distribuzione normale standard. La statistica test che conteggia il numero di dati, tra gli n considerati, che in valore assoluto eccedono $r\sigma_X$ è dunque definita da $C_{r,n} = \sum_{k=1}^n Z_k$ e sotto ipotesi nulla segue una distribuzione binomiale:

$$\Pr(C_{r,n} = c | r, n) = \binom{n}{c} p_r^c (1 - p_r)^{n-c}. \quad (2.4)$$

I pedici indicano come la statistica sia definita dalle scelte fatte per i due parametri r ed n .

2.1.2 *Norm Filter.*

La sovrapposizione di un segnale alla usuale presenza del rumore nella serie storica porterà ad un locale incremento dell'energia nella serie stessa, ovvero

ad un aumento del massimo della funzione di autocorrelazione che si ottiene in ogni caso per il *lag* 0 (cioè la varianza). Perciò in [4] si propone di utilizzare come statistica test l'energia di n dati consecutivi nella serie storica data da $E_n = \sum_{k=1}^n Y_k^2$. Supponendo che σ_X^2 sia nota, E_n/σ_X^2 sotto ipotesi nulla ha una distribuzione chi quadro con n gradi di libertà, mentre sotto ipotesi alternativa ha una distribuzione chi quadro non centrale con gli stessi gradi di libertà ed un parametro di non centralità che dipende dal segnale. L'unico parametro da cui dipende questa statistica è la cardinalità n dell'insieme di dati consecutivi considerato.

2.1.3 Norm of the Autocorrelation

La funzione di autocorrelazione per gli n dati presi in esame è data da:

$$A_l = \frac{1}{n-l} \sum_{k=1}^{n-l} Y_k Y_{k+l}. \quad (2.5)$$

per $l = 0, 1, \dots, n-1$. Sotto ipotesi nulla, se σ_X^2 è nota, A_l/σ_X^2 non è altro che una possibile stima della correlazione del rumore al *lag* l . Dal momento che si assume di avere a che fare con un rumore bianco, cioè non correlato, il suo contributo ad A_l si concentrerà nel *lag* $l = 0$. Viceversa il segnale che si sta cercando sarà sicuramente correlato e contribuirà ad A_l anche per altri *lags* l . Sulla base di queste considerazioni in [9] viene definita una norma della funzione di autocorrelazione in cui si esclude il contributo al *lag* $l = 0$:

$$\|A\|_n = \sqrt{\frac{1}{n-1} \sum_{l=1}^{n-1} A_l^2}, \quad (2.6)$$

e la si usa come statistica test per individuare la presenza di un segnale negli n dati considerati. La distribuzione esatta per questa statistica non è triviale anche nel caso dell'ipotesi nulla e viene determinata con simulazioni o con approssimazioni ad altre distribuzioni.

2.1.4 Slope Filter

Con le assunzioni fatte sulle caratteristiche del rumore non ci si aspetta che esso manifesti alcun tipo di *trend* su n dati consecutivi. Perciò in un adattamento

lineare, col metodo dei minimi quadrati, la stima del coefficiente di pendenza dovrebbe dare valori prossimi a zero a meno che un segnale non sia sovrapposto al rumore. Questo coefficiente può essere riscritto come:

$$S_n = \frac{1}{n} \sum_{k=1}^n \frac{nk - \sum_{i=1}^n i}{\sum_{i=1}^n i^2 - (\sum_{i=1}^n i)^2} Y_k, \quad (2.7)$$

e in [9] viene utilizzato come statistica test per verificare la sovrapposizione di un segnale al rumore. Nel caso in cui tale segnale sia assente, la distribuzione di S_n è gaussiana (in quanto combinazione lineare di variabili gaussiane) con media nulla e varianza pari a $12/[n(n^2 - 1)]$; altrimenti la distribuzione rimane ancora gaussiana e con la medesima varianza, ma con una media data dalla pendenza media del segnale nella finestra di dati considerata. La larghezza n di tale finestra è l'unico parametro da cui dipende la statistica S_n .

2.1.5 *Offset Filter*

In presenza di un segnale le distribuzioni marginali dei singoli dati Y_k continuano a rimanere gaussiane e con la stessa varianza, ma con un valore della media in generale non nullo e diverso per ogni singolo dato. Se il segnale è caratterizzato da variazioni temporali lente rispetto al tempo di campionamento, allora i valori delle medie per dati consecutivi sono tra loro simili e si può pensare che quindi anche la media $O_n = \sum_{k=1}^n Y_k/n$ si discosti da zero per un qualche insieme di n dati consecutivi. Anche in questo caso la distribuzione della statistica test O_n è gaussiana con varianza pari a σ_X^2/n e media o nulla, nel caso in cui il segnale sia assente, oppure pari alla media del segnale nella finestra di dati considerata, la cui larghezza n è l'unico parametro da scegliere anche per questo tipo di statistica, presentata in [28].

2.1.6 *Peak Correlator*

Un metodo per cogliere la presenza di un segnale sommerso dal rumore è ottenibile dalla ricerca del suo picco principale. A tale scopo in [5] si propone di calcolare la correlazione tra la sequenza di dati ed un andamento di tipo gaussiano $g_\tau(k) = \exp(-k^2/\tau^2)$, definendo come statistica test $P_{\tau,n} = \sum_{k=1}^n g_\tau(k - n/2) Y_k$. Ancora una volta sotto ipotesi nulla la statistica test è combinazione lineare di variabili gaussiane a media nulla e dunque

ha a sua volta una distribuzione gaussiana a media nulla la cui varianza è data dal prodotto di σ_X^2 per la media dei quadrati dei pesi impiegati nella combinazione lineare. Nel caso in cui valga l'ipotesi alternativa la distribuzione rimane gaussiana e con la stessa varianza, ma il valore medio è dato dalla correlazione tra il segnale puro e g_τ . È da notare come questa statistica sia prevedibilmente potente solo nel caso in cui la scelta del parametro τ sia operata sulla base di una buona conoscenza della larghezza del picco principale del segnale che deve essere rivelato. In mancanza di questa conoscenza sul segnale è necessario approntare una griglia di valori per τ e con essa definire un insieme di statistiche simultanee da impiegare per la rivelazione del segnale stesso.

2.2 I test wavelet

Tutte le statistiche introdotte in precedenza sono trasformazioni della serie storica Y_1, Y_2, \dots, Y_n assai poco influenzate dalla struttura di correlazione dei dati ai lags più piccoli, zero escluso. In altre parole sono principalmente i contributi di bassa e media frequenza della serie a determinare il valore delle statistiche test. Ciò ha una sua logica, dal momento che il teorema di Nyquist (si veda [40]) indica come adeguata al campionamento di un segnale una frequenza almeno doppia rispetto alla frequenza più alta del segnale stesso. Tipicamente si adottano campionamenti ancora più fitti per avere una migliore descrizione del segnale. In effetti anche nel caso di VIRGO verrà adottata una frequenza di campionamento di 10-20 KHz, molto maggiore rispetto alle frequenze più elevate che dovrebbero caratterizzare i segnali da rivelare.

L'affermazione appena fatta svela come in realtà non si sia proprio del tutto privi di indicazioni sui segnali da rivelare, ma che, anche nei casi peggiori, si abbia un'idea grossolana della banda di frequenza ad essi associata, grazie ai modelli teorici proposti dai fisici. Inoltre si assume che tali modelli forniscano un'indicazione almeno approssimativa sulle durate dei segnali da rivelare. Queste conoscenze permettono di fare scelte sui valori da adottare per i parametri delle statistiche descritte in precedenza o per costruirne di nuove.

Risulta dunque logico cercare di costruire nuove statistiche pensandole come filtri passa-basso o passa-banda da applicare alla porzione di serie storica di volta in volta sotto esame.

Da questo genere di considerazioni nasce l'idea di utilizzare lo strumento delle trasformate wavelet per costruire nuovi test di ipotesi finalizzati alla soluzione del problema della rivelazione di segnali transienti di forma ignota. Infatti con le wavelets è possibile definire un insieme ben strutturato di filtri passa-banda e passa-basso con cui ottenere una scomposizione del segnale nel dominio tempo-frequenza. Ciò vuol dire che si ha la possibilità di seguire l'evoluzione temporale della rilevanza dei contributi al segnale che provengono dalle varie bande di frequenza. Questo non sarebbe possibile con un'usuale trasformata di Fourier, in quanto gli elementi della base di questa trasformata, seni e coseni, sono definiti sull'intero asse reale senza essere localizzate attorno ad un suo punto specifico.

La trasformata wavelet discreta utilizza un insieme minimale di elementi per descrivere il segnale senza alcuna perdita di informazione. Ammesso infatti che tale segnale sia discreto ed appartenga allo spazio $\ell^2(\mathbb{Z})$ delle successioni in \mathbb{Z} a quadrati sommabili, esso viene scomposto su una base completa e linearmente indipendente di tale spazio. Con altre trasformate wavelet, come la continua o la discreta non decimata, la scomposizione avviene invece su una base sovracompleta, nella quale il numero di componenti è sensibilmente maggiore. Per ulteriori dettagli si rimanda all'Appendice A.

Consideriamo dunque la serie storica di dati originati da Y_k , $k \in \mathbb{Z}$, ed applichiamo ad essa una trasformata wavelet discreta. Si ottengono i coefficienti:

$$d_{j,m}^{(Y)} = \langle Y | \psi_{j,m} \rangle = \frac{1}{\sqrt{2^j}} \sum_{k \in \mathbb{Z}} Y_k \bar{\psi} \left(\frac{k - 2^j m}{2^j} \right), \quad (2.8)$$

dove ψ è una funzione con determinate proprietà che agisce come un filtro passa-banda ed è chiamata col nome di *wavelet madre*; da essa infatti discendono tutte le altre wavelets della base su cui si scompone la serie Y_k :

$$\psi_{j,m}(k) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{k - 2^j m}{2^j} \right), \quad (2.9)$$

ottenute con dilatazioni e traslazioni dalla wavelet madre. Nel caso della trasformata discreta il parametro che controlla la dilatazione, j , e quello che regola la localizzazione temporale, m , assumono soltanto valori interi (positivi nel caso di j).

Come conseguenza di questa impostazione si ha che la risposta in frequenza delle funzioni $\psi_{j,m}$ risulta "compressa" di un fattore 2 nel passare da j a $j+1$.

A j viene dato il nome di parametro di scala, in quanto ad esso è legata la scala temporale delle oscillazioni nei dati che vengono catturate dai coefficienti $d_{j,m}$. Ad m invece si dà il nome di parametro di localizzazione, in quanto è per suo mezzo che si distinguono contributi relativi a tempi diversi, anche se caratterizzati dalla stessa scala. È da notare tuttavia che un dato valore di m si riferisce ad istanti diversi per scale j diverse: infatti la cardinalità dell'insieme di coefficienti che discendono da un intervallo temporale finito di dati dimezza ad ogni scala.

Questo indica che per le scale più piccole, ossia per le frequenze maggiori, la localizzazione temporale è migliore rispetto alle scale più larghe, grazie ad un maggior numero di parametri di localizzazione in corrispondenza di uno stesso intervallo di dati nella serie storica oggetto di studio. Viceversa per le scale più larghe, cioè alle basse frequenze, si ha una migliore risoluzione in frequenza, come conseguenza della “compressione” della larghezza di banda delle $\psi_{j,m}$. Ciò che rimane costante, per tutte le wavelets della base, è il rapporto tra frequenza caratteristica e larghezza di banda.

Oltre alla famiglia di filtri passa-banda appena descritti, alla wavelet madre ψ è collegata anche una famiglia di filtri passa-basso che risulta utile per i nostri scopi. Infatti a ψ si può univocamente associare una funzione ϕ con le caratteristiche di un filtro passa-basso che prende il nome di *funzione di scala* o *wavelet padre*, sebbene non abbia affatto le caratteristiche di una wavelet. Da essa, con una relazione analoga alla 2.9, si ottiene una famiglia di versioni dilatate e traslate $\phi_{j,m}$: i contributi alla serie storica Y_k per scale più larghe della j -esima vengono catturati dai coefficienti:

$$c_{j,m}^{(Y)} = \langle y | \phi_{j,m} \rangle = \frac{1}{\sqrt{2^j}} \sum_{k \in \mathbb{Z}} Y_k \bar{\phi} \left(\frac{k - 2^j m}{2^j} \right). \quad (2.10)$$

Perciò è possibile scomporre la serie storica Y_k nei coefficienti $d_{j,m}^{(Y)}$ per esaminare la rilevanza dei contributi delle scale da 1 a J , mentre per l'insieme delle scale maggiori di J si fa ricorso ai coefficienti $c_{J,m}^{(Y)}$, senza perdere in tal modo alcuna informazione sulla serie storica originaria; infatti tale serie può essere ricostruita esattamente a partire da questo insieme di coefficienti, grazie ad una trasformata inversa, [10].

Assumiamo adesso di essere interessati all'analisi della serie storica Y_k per verificare se in essa ad un rumore bianco gaussiano sia sovrapposto qualche

tipo di segnale transiente s_k nell'intervallo temporale $1 \leq k \leq n = 2^l$; supponiamo inoltre di avere qualche motivazione che ci induca a ritenere tale segnale principalmente caratterizzato nelle sue oscillazioni da una scala temporale dell'ordine di $2^j \cdot t_c$, per $2 < j < l$, dove con t_c si indica il tempo di campionamento fisso che intercorre tra due consecutivi valori di k . In questo caso, sviluppando un'idea suggerita da Carmona in [8], si propone di utilizzare i coefficienti wavelet introdotti in precedenza per costruire la seguente statistica test:

$$T_{\psi_j, n} = \frac{\sum_{m=1}^{2^{l-j}} |d_{j,m}^{(Y)}|^2}{\sum_{m=1}^{2^{l-1}} |d_{1,m}^{(Y)}|^2 + \sum_{m=1}^{2^{l-2}} |d_{2,m}^{(Y)}|^2}. \quad (2.11)$$

I pedici ψ_j ed n sottolineano la dipendenza di questa statistica dalle scelte fatte per la wavelet madre ψ da usare, per la scala j e la larghezza $n = 2^l$ dell'intervallo di dati da analizzare. Infatti nelle sommatorie che compaiono nella precedente espressione gli intervalli di valori per il parametro di localizzazione m sono scelti in modo da considerare solo i coefficienti wavelet $d_{\cdot, m}$ che discendono dalla restrizione a Y_1, Y_2, \dots, Y_n dell'intera serie storica. È bene rimarcare che è proprio grazie ai parametri ψ , j ed n che è possibile trarre profitto, in termini di potenza della statistica test, dalle più o meno vaghe informazioni a priori di cui si dispone per l'andamento, la banda di frequenza e la durata del segnale da rivelare.

L'idea alla base dell'introduzione della statistica $T_{\psi_j, n}$ è che il valore atteso dei termini $|d_{\cdot, m}^{(Y)}|^2$, in corrispondenza di un intervallo limitato di valori di m , sia maggiore nel caso in cui valga l'ipotesi alternativa H_1 del modello 2.1, e che ciò sia tanto più evidente quanto maggiore è la rilevanza della scala j per il segnale da rivelare.

Al denominatore vengono invece utilizzati i quadrati dei coefficienti wavelet alle scale più piccole; in tal modo, supponendo che il segnale dia un contributo irrilevante alla serie storica a queste scale, il denominatore dipende esclusivamente dalle caratteristiche del rumore e svolge il ruolo di adattare la statistica test al valore *locale* della varianza del rumore. In altre parole la statistica $T_{\psi_j, n}$ risulta adeguata anche nel caso in cui il rumore previsto nel modello 2.1, pur rimanendo non correlato, manifesti un'evoluzione temporale, non nota, della

propria varianza σ_X^2 , a patto che tale evoluzione sia caratterizzata da tempi più lunghi rispetto all'intervallo di n dati analizzato.

D'altra parte l'ipotesi che il segnale abbia scarso rilievo alle scale 1 e 2 è del tutto ragionevole se la frequenza di campionamento f_c risulta adeguata per il segnale stesso, come si deduce dalle considerazioni svolte in precedenza sul teorema di Nyquist. Eventualmente si può modificare il denominatore di $T_{\psi_j, n}$, considerando soltanto i coefficienti alla scala più fine, nel caso in cui non si sia certi di poter trascurare gli effetti del segnale alla scala $j = 2$.

Sfruttando una serie di ben note proprietà della trasformata wavelet discreta è possibile arrivare ad una formalizzazione più rigorosa delle caratteristiche intuitivamente attribuite alla statistica $T_{\psi_j, n}$. Vediamo come ciò sia possibile.

Proposizione 2.1 *La trasformata wavelet discreta è lineare.*

Dim: Se la serie di dati Y_k è ottenuta come combinazione lineare di altre due serie storiche U_k e V_k :

$$Y_k = a_u U_k + a_v V_k, \quad (2.12)$$

allora la sua trasformata wavelet discreta è ottenuta come combinazione lineare delle trasformate delle due componenti:

$$\begin{aligned} d_{j,m}^{(Y)} &= \sum_{k \in \mathbb{Z}} \{a_u U_k + a_v V_k\} \bar{\psi}_{j,m}(k) \\ &= a_u \sum_{k \in \mathbb{Z}} U_k \bar{\psi}_{j,m}(k) + a_v \sum_{k \in \mathbb{Z}} V_k \bar{\psi}_{j,m}(k) \\ &= a_u d_{j,m}^{(U)} + a_v d_{j,m}^{(V)}. \end{aligned} \quad (2.13)$$

Analoga dimostrazione vale per i coefficienti di scala. *qed*

Proposizione 2.2 *Se la serie di dati X_k non è correlata, ha valore atteso nullo e viene scomposta con una trasformata wavelet discreta ed ortogonale allora anche i coefficienti wavelet $d_{j,m}^{(X)}$ e di scala $c_{j,m}^{(X)}$ risultano non correlati tra loro, con valore atteso nullo e la stessa varianza σ_X^2 dei dati.*

Dim: In una trasformata wavelet discreta ortonormale la wavelet madre ψ è tale che, per ogni $(j, j') \in \mathbb{N} \times \mathbb{N}$ e $(m, m') \in \mathbb{Z} \times \mathbb{Z}$:

$$\langle \psi_{j', m'} | \psi_{j, m} \rangle = \sum_{k \in \mathbb{Z}} \psi_{j', m'}(k) \bar{\psi}_{j, m}(k) = \delta_{j j'} \delta_{m m'}, \quad (2.14)$$

L'ipotesi di non correlazione dei dati X_k implica che per ogni $(k, k') \in \mathbb{Z} \times \mathbb{Z}$:

$$\mathcal{E} [X_k \overline{X_{k'}}] = \delta_{kk'} \sigma_X^2 \quad (2.15)$$

Da queste relazioni, sfruttando la linearità del valore atteso, si ottiene:

$$\begin{aligned} \mathcal{E} [d_{j,m}^{(X)} \overline{d_{j',m'}^{(X)}}] &= \mathcal{E} \left[\sum_{k \in \mathbb{Z}} X_k \overline{\psi_{j,m}(k)} \sum_{k' \in \mathbb{Z}} \overline{X_{k'}} \psi_{j',m'}(k') \right] \\ &= \sum_{k \in \mathbb{Z}} \sum_{k' \in \mathbb{Z}} \psi_{j',m'}(k') \overline{\psi_{j,m}(k)} \mathcal{E} [X_k \overline{X_{k'}}] \\ &= \sigma_X^2 \sum_{k \in \mathbb{Z}} \psi_{j',m'}(k) \overline{\psi_{j,m}(k)} \\ &= \delta_{jj'} \delta_{mm'} \sigma_X^2. \end{aligned} \quad (2.16)$$

Dunque la correlazione attesa per coefficienti wavelet distinti è nulla se la serie storica di partenza non è correlata e la trasformata wavelet discreta è ortogonale. La varianza dei coefficienti è la stessa dei dati. Per i coefficienti di scala la dimostrazione è analoga. qed

Proposizione 2.3 *Se i dati Y_k hanno distribuzione di probabilità normale, allora anche i coefficienti wavelet e di scala sono normalmente distribuiti.*

Dim: I coefficienti sono combinazioni lineari dei dati, come si vede dalle equazioni 2.8 e 2.10. Ma combinazioni lineari di variabili aleatorie normali hanno a loro volta una distribuzione di probabilità normale. qed

Data la linearità della trasformata wavelet, facendo riferimento al modello 2.1, si ha che:

$$d_{j,m}^{(Y)} = \begin{cases} d_{j,m}^{(X)}, & \text{sotto } H_0; \\ d_{j,m}^{(X)} + d_{j,m}^{(s)}, & \text{sotto } H_1, \end{cases} \quad (2.17)$$

e dunque, per i moduli quadri dei coefficienti:

$$\left| d_{j,m}^{(Y)} \right|^2 = \begin{cases} \left| d_{j,m}^{(X)} \right|^2, & \text{sotto } H_0; \\ \left| d_{j,m}^{(X)} \right|^2 + \left| d_{j,m}^{(s)} \right|^2 + d_{j,m}^{(s)} \overline{d_{j,m}^{(X)}} + \\ \quad + d_{j,m}^{(X)} \overline{d_{j,m}^{(s)}}, & \text{sotto } H_1. \end{cases} \quad (2.18)$$

Se la serie di rumore X_k non è correlata e con valore atteso nullo, con un'opportuna scelta della wavelet madre, si possono sfruttare i risultati della proprietà 2.2 per dire che:

$$\mathcal{E} [d_{j,m}^{(X)}] = 0, \quad (2.19)$$

$$\mathcal{E} \left[\left| d_{j,m}^{(X)} \right|^2 \right] = \sigma_Y^2, \quad (2.20)$$

da cui si ottiene, considerando il valore atteso di entrambi i membri dell'equazione 2.18:

$$\mathcal{E} \left[\left| d_{j,m}^{(Y)} \right|^2 \right] = \begin{cases} \sigma_X^2 & \text{sotto } H_0; \\ \sigma_X^2 + \left| d_{j,m}^{(s)} \right|^2 & \text{sotto } H_1. \end{cases} \quad (2.21)$$

Perciò la presenza del segnale ha l'effetto di aumentare il valore atteso del modulo quadro dei coefficienti wavelet. La statistica test $T_{\psi_j, n}$ definita in 2.11 viene costruita, sfruttando le vaghe conoscenze a priori sul segnale, in modo tale che il contributo del segnale $\left| d_{j,m}^{(s)} \right|^2$ sia grande per i coefficienti al numeratore e trascurabile per quelli al denominatore. Sotto ipotesi alternativa dunque i valori assunti dalla statistica test risulteranno maggiori rispetto al caso dell'ipotesi nulla. Inoltre i coefficienti utilizzati nelle somme a denominatore, se effettivamente sono influenzati in misura trascurabile dalla eventuale presenza del segnale, hanno il valore atteso del modulo quadro uguale alla varianza locale del rumore stesso e possono quindi essere utilizzati come stimatori di tale varianza.

A fronte dei vantaggi appena illustrati la scelta di una singola scala j , da usare nella definizione della statistica $T_{\psi_j, n}$, in taluni casi può risultare una limitazione pericolosa. Infatti non è sempre detto che siano affidabili le indicazioni a priori sulle caratteristiche del segnale che ci inducono a scegliere un determinato valore per j : ne consegue una perdita di robustezza del test di ipotesi. Inoltre, anche se la scala j è effettivamente quella di maggior rilievo per il segnale da rivelare, è possibile che anche i contributi ad altre scale siano cospicui e che dunque possano essere utilizzati per la definizione di statistiche più potenti di quella discussa fino a questo punto.

Sulla base di questo genere di considerazioni nel seguito vengono proposte alcune originali generalizzazioni del tipo di statistica definita dall'equazione 2.11.

In primo luogo consideriamo il caso in cui si ritenga di dover cercare il contributo del segnale in un insieme di scale $\mathcal{J} \subset \{j \in \mathbb{N} : 2 < j < \log_2(n)\}$, dove n è il numero di dati presi in considerazione. In pratica si verifica comunemente che l'insieme \mathcal{J} sia costituito da scale consecutive, ad esempio quando ci si aspetti di trovare il segnale su un'ampia banda di frequenze. In ogni caso con

la notazione adottata non si perde in generalità. Per ognuna delle scale $j \in \mathcal{J}$ l'equazione 2.11 definisce una statistica test $T_{\psi_j, n}$. Un approccio possibile è quello di sfruttare, per la costruzione di un test di ipotesi, l'inferenza che si può ottenere da tutte queste statistiche considerate come simultanee; si veda in proposito [35]. Ma la soluzione più semplice è sicuramente quella di cercare di combinare le varie $T_{\psi_j, n}$ in un'unica statistica test. In questo senso risulta intuitivo e attraente considerare il massimo valore assunto dalle statistiche $T_{\psi_j, n}$:

$$T_{\max, \psi_{\mathcal{J}}, n} = \max_{j \in \mathcal{J}} (T_{\psi_j, n}). \quad (2.22)$$

Tuttavia il fatto che le $T_{\psi_j, n}$ abbiano distribuzioni diverse e siano tra loro dipendenti a causa del comune denominatore, rende piuttosto complessa sia l'inferenza simultanea che lo studio analitico della distribuzione di $T_{\max, \mathcal{J}}$. Ma proprio la presenza di un comune denominatore depone a favore di un'altra statistica test, ottenuta con una combinazione lineare delle $T_{\psi_j, n}$, come ad esempio in:

$$T_{\Sigma, \psi_{\mathcal{J}}, n} = \sum_{j \in \mathcal{J}} T_{\psi_j, n}. \quad (2.23)$$

È chiaro che questa statistica risulta particolarmente appropriata al caso in cui la rilevanza segnale sia confrontabile per tutte le scale $j \in \mathcal{J}$. Minori benefici si ottengono invece se la statistica $T_{\Sigma, \psi_{\mathcal{J}}, n}$ viene adottata perchè si sa che il segnale ha un contributo preponderante ad una sola scala $j \in \mathcal{J}$, ma non si è in grado di specificare quale; in quest'ultimo caso la statistica $T_{\max, \psi_{\mathcal{J}}, n}$ risulterebbe verosimilmente di maggiore potenza.

Le statistiche wavelet descritte fino a questo punto hanno tutte le caratteristiche di filtri passa-banda; tuttavia può essere utile, ai fini della rivelazione di segnali transienti, considerare anche statistiche test con le caratteristiche di un filtro passa-basso. In termini di scale ciò vuol dire cercare segnali su tutte le scale più grossolane a partire da una prefissata che indichiamo con J . D'altra parte se le funzioni wavelet hanno le caratteristiche di un filtro passa-banda, le funzioni di scala ad esse associate sono proprio dei filtri passa-basso. Da queste considerazioni nasce lo spunto per la formulazione di una originale statistica test da utilizzare nel contesto della rivelazione di segnali transienti: nel numeratore dell'espressione in 2.11 si propone di sostituire i coefficienti wavelet con i coefficienti di scala che colgono nei dati strutture su tutte le scale più larghe

della J -esima:

$$T_{\phi_J, n} = \frac{\sum_{m=1}^{2^{l-J}} |c_{J,m}^{(Y)}|^2}{\sum_{m=1}^{2^{l-1}} |d_{1,m}^{(Y)}|^2 + \sum_{m=1}^{2^{l-2}} |d_{2,m}^{(Y)}|^2}. \quad (2.24)$$

Si fa notare che il denominatore rimane lo stesso dell'espressione 2.11 e mantiene la stessa funzione, discussa in precedenza, di adattamento al valore locale della varianza del rumore.

Il numeratore con i coefficienti di scala $c_{J,m}$ conferisce alla statistica $T_{\phi_J, n}$ caratteristiche strutturalmente diverse rispetto alle statistiche definite con soli coefficienti wavelet. Infatti i coefficienti di scala, a differenza di quelli wavelet, sono influenzati anche dagli scostamenti della serie dei dati dal valore medio nullo; d'altra parte proprio questi scostamenti erano stati sfruttati allo scopo di rivelare l'eventuale presenza di un segnale nella statistica *offset filter* descritta in precedenza.

Inoltre non si pone il problema di combinare tra loro diverse statistiche $T_{\phi_J, n}$ per diversi valori di J , in analogia a quanto discusso per le $T_{\psi_j, n}$. Infatti se $J_1 < J_2$ la statistica $T_{\phi_{J_1}, n}$ coglie anche i contributi delle scale coperte da $T_{\phi_{J_2}, n}$ ($j > J_2$) oltre a quelli corrispondenti a scale comprese tra J_1 e J_2 . È chiaro quindi che, anche sotto ipotesi nulla, le statistiche $T_{\phi_{J_1}, n}$ e $T_{\phi_{J_2}, n}$ risultano tra loro dipendenti non solo a causa del comune denominatore, ma anche a causa dei rispettivi numeratori, correlati tra loro.

Risulta invece di qualche interesse cercare di combinare la statistica $T_{\phi_J, n}$ con la statistica $T_{\Sigma, \psi_{\mathcal{J}}, n}$, definita in 2.23, per $\mathcal{J} \subset \{j \in \mathbb{N} : 2 < j \leq J\}$. Anche in questo caso il denominatore comune favorisce la soluzione di una semplice somma delle due statistiche considerate:

$$T_{\psi_{\mathcal{J}}, \phi_J, n} = T_{\Sigma, \psi_{\mathcal{J}}, n} + T_{\phi_J, n}. \quad (2.25)$$

Naturalmente la statistica appena definita risulta particolarmente appropriata per segnali che siano rilevanti sia per scale $j \in \mathcal{J}$ che per quelle $j > J$. Tuttavia nella pratica comune è piuttosto insolito dover considerare scale non contigue tra loro.

Il caso particolare dell'espressione 2.25 che si ottiene quando $\mathcal{J} = \{J\}$ merita una considerazione particolare. Infatti in questo caso si ottiene:

$$T_{\psi_J, \phi_J, n} = T_{\psi_J, n} + T_{\phi_J, n}, \quad (2.26)$$

una statistica che può essere considerata come alternativa della $T_{\phi_{J-1},n}$, dal momento che entrambe risultano sensibili per le stesse scale: $j > J - 1$, ossia $j > J$ e $j = J$. Questa considerazione empirica trova conferma nella seguente proposizione originale.

Proposizione 2.4 *Se si utilizza una trasformata wavelet discreta ortogonale allora vale la seguente relazione tra le statistiche test definite in 2.11 e 2.24:*

$$T_{\phi_{J-1},n} = T_{\psi_J,n} + T_{\phi_J,n}. \quad (2.27)$$

Dim: Quando la trasformata wavelet è ortogonale, allora la serie originale dei dati può essere ricostruita a partire dai coefficienti $d_{j,m}^{(Y)}$ e $c_{j,m}^{(Y)}$ grazie ad una delle relazioni (si veda ad esempio [31]):

$$Y_k = \sum_{j \in \mathbb{N}^+} \sum_{m \in \mathbb{Z}} d_{j,m}^{(Y)} \psi_{j,m}(k) \quad (2.28)$$

$$Y_k = \sum_{j=1}^J \sum_{m \in \mathbb{Z}} d_{j,m}^{(Y)} \psi_{j,m}(k) + \sum_{m \in \mathbb{Z}} c_{J,m}^{(Y)} \phi_{J,m}(k). \quad (2.29)$$

Eguagliando i secondi membri delle precedenti espressioni consegue che, qualsiasi sia la serie Y_k e $\forall J > 0$:

$$\sum_{j=J+1}^{+\infty} \sum_{m \in \mathbb{Z}} d_{j,m}^{(Y)} \psi_{j,m}(k) = \sum_{m \in \mathbb{Z}} c_{J,m}^{(Y)} \phi_{J,m}(k). \quad (2.30)$$

Considerando il modulo quadro di entrambi i membri e tenendo conto delle ipotesi di ortogonalità sulle funzioni wavelet e di scala, con semplici manipolazioni si ottiene:

$$\sum_{j=J+1}^{+\infty} \sum_{m \in \mathbb{Z}} \left| d_{j,m}^{(Y)} \right|^2 = \sum_{m \in \mathbb{Z}} \left| c_{J,m}^{(Y)} \right|^2 \quad (2.31)$$

Ma la relazione appena scritta vale per ogni $J > 0$ e quindi sostituendo $J - 1$ a J si ottiene:

$$\sum_{j=J}^{+\infty} \sum_{m \in \mathbb{Z}} \left| d_{j,m}^{(Y)} \right|^2 = \sum_{m \in \mathbb{Z}} \left| c_{J-1,m}^{(Y)} \right|^2 \quad (2.32)$$

Sottraendo membro a membro la 2.31 dalla 2.32 e ricombinando i termini ottenuti, si ha:

$$\sum_{m \in \mathbb{Z}} \left| c_{J-1,m}^{(Y)} \right|^2 = \sum_{m \in \mathbb{Z}} \left| d_{J,m}^{(Y)} \right|^2 + \sum_{m \in \mathbb{Z}} \left| c_{J,m}^{(Y)} \right|^2 \quad (2.33)$$

Dal momento che i coefficienti $d_{J,m}^{(Y)}$ si possono ottenere dalla trasformata wavelet discreta dei coefficienti $c_{J-1,m}^{(Y)}$, come dimostrato in [34], la relazione precedente continua a valere anche restringendo le tre sommatorie ai soli valori di m collegati agli indici $k \in [1, n = 2^l]$ della serie $Y_k, \forall k \in \mathbb{Z}$. Perciò per $J \geq l$ si ha:

$$\sum_{m=1}^{2^{l-(J-1)}} \left| c_{J-1,m}^{(Y)} \right|^2 = \sum_{m=1}^{2^{l-J}} \left| d_{J,m}^{(Y)} \right|^2 + \sum_{m=1}^{2^{l-J}} \left| c_{J,m}^{(Y)} \right|^2 \quad (2.34)$$

Dividendo ambo i membri per:

$$\sum_{m=1}^{2^{l-1}} \left| d_{1,m}^{(Y)} \right|^2 + \sum_{m=1}^{2^{l-2}} \left| d_{2,m}^{(Y)} \right|^2,$$

si ottiene proprio la 2.27. *qed*

2.2.1 Distribuzioni di probabilità delle statistiche test

Nel caso in cui la serie storica del rumore X_k nel modello 2.1 sia non correlata, gaussiana e con variabili identicamente distribuite, è possibile individuare analiticamente le distribuzioni di probabilità delle principali statistiche presentate nella precedente sezione, sia sotto l'ipotesi nulla H_0 di solo rumore che sotto ipotesi alternativa H_1 di sovrapposizione di qualche tipo di segnale.

Con l'eccezione della statistica $T_{\max, \psi_{\mathcal{J}}, n}$, tutte le altre discusse hanno una struttura del tipo:

$$T = \frac{\sum_{i \in \mathcal{I}_1} \left| d_i^{(Y)} \right|^2}{\sum_{i \in \mathcal{I}_2} \left| d_i^{(Y)} \right|^2} \quad (2.35)$$

dove \mathcal{I}_1 e \mathcal{I}_2 sono insiemi disgiunti, $\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset$, di coppie di indici $(j, m)_i$, mentre le variabili $d_i^{(Y)}$ sono coefficienti che derivano dalla scomposizione wavelet di n dati consecutivi Y_k della serie storica. Indichiamo con $\nu_1 = |\mathcal{I}_1|$ e $\nu_2 = |\mathcal{I}_2|$ le cardinalità degli insiemi \mathcal{I}_1 e \mathcal{I}_2 rispettivamente.

Proposizione 2.5 *Se vale il modello 2.1 con X_k indipendenti, a media nulla, identicamente e normalmente distribuite, e se il contributo del segnale è trascurabile nel denominatore delle statistiche T definite in 2.35, allora:*

$$T \frac{\nu_2}{\nu_1} \sim \begin{cases} F_{\nu_1, \nu_2} & \text{sotto } H_0; \\ F_{\nu_1, \nu_2, \delta^{(s)}} & \text{sotto } H_1. \end{cases} \quad (2.36)$$

dove F_{ν_1, ν_2} è una distribuzione di Fisher con ν_1 e ν_2 gradi di libertà, mentre $F_{\nu_1, \nu_2, \delta^{(s)}}$ è una distribuzione di Fisher non centrale con gli stessi gradi di libertà di F_{ν_1, ν_2} . Il parametro di non centralità $\delta^{(s)}$ dipende solo dalle caratteristiche del segnale ed è pari al valore che assumerebbe, in assenza di rumore, il numeratore del secondo membro dell'espressione 2.35.

Dim: Data la linearità sia del modello 2.1 che della trasformata wavelet discreta, come risulta dalla proposizione 2.1, si ha:

$$d_i^{(Y)} = \begin{cases} d_i^{(X)}, & \text{sotto } H_0; \\ d_i^{(X)} + d_i^{(s)}, & \text{sotto } H_1, \end{cases} \quad (2.37)$$

dove $d_i^{(X)}$ e $d_i^{(s)}$ sono i coefficienti corrispondenti di $d_i^{(Y)}$ per la serie del solo rumore X_k e per quella del solo segnale s_k rispettivamente. Per la proposizione 2.3 si ha che le variabili aleatorie $d_i^{(X)}$, $i \in \mathcal{I}_1 \cup \mathcal{I}_2$, hanno distribuzione normale e centrale, mentre, essendo $\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset$, dalla proposizione 2.2 consegue che le $d_i^{(X)}$ sono tra loro non correlate, a media nulla e con varianza σ_X^2 . Essendo gaussiane e non correlate le variabili $d_i^{(X)}$ sono indipendenti tra loro. L'aggiunta a $d_i^{(X)}$ del termine $d_i^{(s)}$ ha il solo effetto di spostare la media delle variabili $d_i^{(Y)}$, che tuttavia, condizionatamente alla conoscenza del segnale s_k , rimangono gaussiane, con la stessa varianza σ_X^2 , non correlate e dunque indipendenti tra loro. Sono dunque indipendenti anche le variabili:

$$U_i = \frac{d_i^{(Y)}}{\sigma_X} \sim \begin{cases} N_{0,1}, & \text{sotto } H_0; \\ N_{\mu_i^{(s)},1}, & \text{sotto } H_1, \end{cases} \quad (2.38)$$

dove $N_{0,1}$ è una distribuzione normale standard, mentre $N_{\mu_i^{(s)},1}$ è una distribuzione normale di varianza unitaria e media $\mu_i^{(s)} = d_i^{(s)}/\sigma_X$.

Il modulo quadro di una variabile distribuita secondo una normale standard ha distribuzione chi-quadrato con un solo grado di libertà χ_1^2 ; ma nel caso in cui la variabile normale di varianza unitaria abbia media $\mu \neq 0$, la distribuzione

del suo modulo quadro risulterà una chi-quadrato non centrale χ_{1,μ^2}^2 con un grado di libertà e parametro di non centralità μ^2 . Dunque:

$$V_i = |U_i|^2 \sim \begin{cases} \chi_1^2, & \text{sotto } H_0; \\ \chi_{1,\delta_i^{(s)}}^2, & \text{sotto } H_1. \end{cases} \quad (2.39)$$

dove $\delta_i^{(s)} = |d_i(s)|^2/\sigma_X^2$. Si dimostra facilmente che anche le variabili V_i risultano indipendenti tra loro. La somma di ν variabili indipendenti distribuite χ_1^2 ha distribuzione chi-quadrato χ_ν^2 con ν gradi di libertà; nel caso in cui le ν variabili sommate abbiano distribuzioni χ_{1,δ_i}^2 , anche la distribuzione della somma è una $\chi_{\nu,\delta}^2$ con ν gradi di libertà, ma con parametro di non centralità $\delta = \sum \delta_i$. Si ha quindi che, per $p \in \{1, 2\}$:

$$W_p = \sum_{i \in \mathcal{I}_p} V_i \sim \begin{cases} \chi_{\nu_p}^2, & \text{sotto } H_0; \\ \chi_{\nu_p,\delta_p^{(s)}}^2, & \text{sotto } H_1, \end{cases} \quad (2.40)$$

dove:

$$\delta_p^{(s)} = \frac{1}{\sigma_X^2} \sum_{i \in \mathcal{I}_p} |d_i^{(s)}|^2. \quad (2.41)$$

Dal momento che $\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset$ si ha che W_1 e W_2 sono combinazioni lineari di due sottoinsiemi disgiunti dell'insieme di variabili indipendenti $\mathcal{V} = \{V_i : i \in \mathcal{I}_1 \cup \mathcal{I}_2\}$; si può dimostrare che ciò comporta l'indipendenza di W_1 e W_2 .

A questo punto si considera che per i coefficienti con indice $i \in \mathcal{I}_2$ si è ipotizzato un contributo del segnale trascurabile rispetto a quello del rumore, e cioè:

$$0 \simeq \delta_2^{(s)} = \frac{1}{\sigma_X^2} \sum_{i \in \mathcal{I}_2} |d_i^{(s)}|^2 \ll \mathcal{E} \left[\frac{1}{\sigma_X^2} \sum_{i \in \mathcal{I}_2} |d_i^{(X)}|^2 \right] = \nu_2. \quad (2.42)$$

Ma in tal caso la distribuzione di W_2 può essere considerata come una chi-quadrato centrale $\chi_{\nu_2}^2$ anche sotto l'ipotesi alternativa H_1 .

Si vede facilmente che l'identità 2.35 può essere riscritta come:

$$T \frac{\nu_2}{\nu_1} = \frac{W_1/\nu_1}{W_2/\nu_2}. \quad (2.43)$$

Il membro di destra è il rapporto di due variabili aleatorie distribuite chi-quadrato, normalizzato dal rapporto dei gradi di libertà: pertanto la sua distribuzione è una F di Fisher con gradi di libertà corrispondenti e con eventuale

parametro di non centralità $\delta^{(s)} = \delta_1^{(s)}$ uguale a quello della distribuzione di W_1 .
Da ciò consegue la 2.36. *qed*

In realtà la statistica $T_{\phi_J, n}$ e tutte quelle da essa derivate non rientrano esattamente nella definizione data dalla 2.35, dal momento che fanno uso di coefficienti di scala oltre che di coefficienti wavelet. Tuttavia per queste statistiche i coefficienti di scala impiegati si riferiscono sempre ad un'unica scala J maggiore di tutte quelle relative ai coefficienti wavelet contestualmente utilizzati: in questa situazione, se la trasformata wavelet è ortogonale, si può ricalcare la dimostrazione della proposizione 2.5 per ottenere un risultato del tutto analogo, con gradi di libertà dati dal numero di coefficienti impiegati a numeratore o a denominatore, senza far distinzione tra quelli di scala e quelli wavelet.

In effetti, grazie all'applicazione ricorsiva della proprietà 2.4, sarebbe anche possibile esprimere la statistica $T_{\phi_J, n}$ usando coefficienti wavelet piuttosto che coefficienti di scala e dunque avvicinandosi ad una struttura come quella definita dalla 2.35:

$$T_{\phi_J, n} = \sum_{j=J+1}^{J'} T_{\psi_j, n} + T_{\phi_{J'}, n}, \quad (2.44)$$

con $J < J' \leq l = \log_2(n)$. Nel membro di destra della 2.44 il numero di coefficienti di scala impiegati è minore che nel membro di sinistra. Tuttavia anche nel caso estremo con $J' = l = \log_2(n)$, nel membro di destra è comunque presente un coefficiente di scala, grazie al quale si mantiene la distinzione strutturale tra le statistiche passa-basso come $T_{\phi_J, n}$ e le statistiche passa-banda tipo $T_{\psi_j, n}$. È però estremamente interessante osservare come in entrambe le formulazioni sono conservati gli stessi numeri ν_1 e ν_2 di termini sommati nei numeratori e nei denominatori (per il denominatore l'osservazione è banale dal momento che rimane sempre lo stesso); d'altra parte se così non fosse, data la proposizione 2.5, avremmo gradi di libertà diversi nelle distribuzioni del primo e del secondo membro della 2.44, un risultato che sarebbe chiaramente assurdo.

La proposizione 2.5 risulta di estrema utilità e verrà più volte sfruttata nel corso di questo lavoro; infatti grazie a tale proposizione è possibile caratterizzare tutte le statistiche test proposte nella sezione 2.2, senza dover ricorrere a lunghe simulazioni. In particolare, per ogni valore α scelto tra 0 e 1 e per ogni statistica T definita come in 2.35, la distribuzione di T sotto ipotesi nulla

consente fissare il valore di soglia $\tau_{\alpha,T}$ tale che $\Pr(T \geq \tau_{\alpha,T}) \leq \alpha$; in tal modo si può definire il seguente test di ipotesi di ampiezza α :

$$\begin{aligned} &\text{si accetti } H_0 \text{ qualora sia } T < \tau_{\alpha,T}, \\ &\text{si accetti } H_1 \text{ qualora sia } T \geq \tau_{\alpha,T}. \end{aligned}$$

La distribuzione di T sotto ipotesi alternativa consente invece di stabilire la potenza π di questo test di ipotesi in presenza di un qualsiasi segnale transiente noto; in pratica basterebbe conoscere il termine $\delta^{(s)}$ che compete a tale segnale per determinare π .

Infine è bene sottolineare che il valore di soglia $\tau_{\alpha,T}$ dipende solamente dalla scelta fatta per α e dai gradi di libertà di T , mentre è completamente indipendente dal valore σ_X^2 della varianza del rumore. Perciò, anche qualora si assista ad un'evoluzione temporale di σ_X^2 , il valore di soglia $\tau_{\alpha,T}$ rimane fissato una volta per tutte dalla scelta di T e di α . Questa desiderabile proprietà è comune a tutte le statistiche presentate nella sezione 2.2, mentre non è condivisa dalle statistiche presentate nella sezione 2.1. La potenza del test π , legata al valore di $\delta^{(s)}$, dipende invece dal valore di σ_X^2 , come risulta evidente dall'equazione 2.41 e come ci si poteva aspettare dal buon senso. Infatti l'intuizione suggerisce che in condizioni di maggiore rumorosità della serie di dati la potenza del test per un dato segnale diminuisca.

2.2.2 Test in intervalli non disgiunti

Tutte le statistiche test presentate per discriminare tra le due ipotesi del modello 2.1 in realtà fanno riferimento soltanto ad un intervallo di n dati della serie storica. D'altra parte un segnale transiente potrebbe essere effettivamente presente nella serie storica pur non manifestandosi nell'intervallo di dati considerato, dal momento che per sua definizione un segnale transiente ha una durata limitata nel tempo. Perciò occorre ripetere il test di ipotesi adottato su una successione di intervalli scelti in modo da coprire l'intera serie storica da analizzare: verrà quindi accettata l'ipotesi alternativa H_1 se il test di ipotesi suggerisce la sua validità in almeno uno di questi intervalli; viceversa per accettare H_0 occorre che il test di ipotesi accetti l'ipotesi nulla in ciascuno degli intervalli della successione. D'altra parte va detto che questo tipo di test non risulta particolarmente interessante se considerato nel suo aspetto globale anziché locale: infatti da un lato la natura stessa dei transienti pone l'accento sui

risultati dei test locali, dall'altro all'aumentare della lunghezza della serie storica è inevitabile che cresca la probabilità di incorrere in un rifiuto dell'ipotesi nulla.

La copertura dell'intera serie storica con il minor numero di intervalli si ottiene affiancando ogni intervallo di n dati al precedente. Tuttavia questa soluzione rischia di compromettere seriamente la potenza del test: infatti il segnale transiente potrebbe distribuire la propria energia su due intervalli consecutivi, con conseguente perdita di potenza in entrambi i test di ipotesi effettuati in ciascuno dei due intervalli. Più in generale si può dire che anche la disposizione degli intervalli di dati da analizzare è un parametro da cui dipende la potenza del test di ipotesi. È chiaro che la potenza massima può essere raggiunta utilizzando intervalli spostati l'uno rispetto all'altro di un solo dato, ma in questa situazione risulta massimo anche il numero di calcoli da effettuare per unità di tempo. Perciò spesso sono vantaggiose soluzioni di compromesso in cui la sovrapposizione tra intervalli è meno estesa, pur mantenendo in termini accettabili la conseguente perdita di potenza del test.

Al di là delle considerazioni qualitative appena svolte, ci si può chiedere quali siano le caratteristiche del test di ipotesi sull'intera serie storica che discendono dal test effettuato nei singoli intervalli e dalla disposizione degli intervalli stessi. Sfruttando un risultato ben noto nel campo dell'inferenza da statistiche simultanee, riportato ad esempio in [35], si ha che il valore atteso del numero $N_{H_1|H_0, n_t}$ di erronei rifiuti di H_0 su n_t test di ampiezza α è dato da:

$$\mathcal{E} [N_{H_1|H_0, n_t}] = n_t \alpha, \quad (2.45)$$

qualsiasi sia la correlazione tra le statistiche degli n_t test effettuati. Perciò, se gli intervalli sono distanziati di s dati l'uno dall'altro e il tempo di campionamento della serie è dato da t_c , avremo un tasso di “falsi allarmi” per unità di tempo pari a:

$$\kappa_{fa} = \lim_{n_t \rightarrow +\infty} \frac{n_t \alpha}{n_t s t_c} = \frac{\alpha}{s t_c}, \quad (2.46)$$

intendendo come falso allarme l'errata rivelazione di un segnale transiente in uno degli intervalli analizzati. La relazione 2.46 è utile per fissare l'ampiezza α dei test di ipotesi sulla base del massimo tasso di falsi allarmi κ_{fa} che si è disposti a tollerare. La scelta di questo valore per κ_{fa} viene fatta sulla base del trattamento specifico di ciascun falso allarme nel contesto in cui si opera.

Ad esempio in un esperimento come VIRGO si può pensare di salvare tutte le sequenze di dati che indicano la presenza di un transiente, per effettuare in un secondo tempo un'analisi più approfondita; perciò il valore di κ_{fa} dovrebbe essere fissato sulla base della capacità di memorizzare un certo numero di sequenze e di utilizzarle effettivamente nell'analisi di secondo livello. D'altra parte un'analisi più approfondita richiederà tempi di calcolo più elevati, per cui occorre che il numero di dati che ad essa vengono sottoposti sia fortemente ridotto rispetto alla serie originaria. Perciò sembrano ragionevoli dei valori di κ_{fa} inferiori ad un falso allarme ogni milione di dati acquisiti che corrisponde comunque a ben 72 falsi allarmi ogni ora per una frequenza di campionamento $f_c = 20$ KHz.

La trattazione dei test in intervalli di dati sovrapposti svolta fino a questo punto, pur essendo formalmente rigorosa, rischia di risultare fuorviante. Infatti in questa trattazione non viene assolutamente considerata la struttura di correlazione tra le statistiche test calcolate nei vari intervalli. Se gli intervalli non sono disgiunti allora, anche sotto l'ipotesi nulla di solo rumore bianco e gaussiano, tali statistiche saranno necessariamente correlate tra loro in quanto calcolate sulla base di insiemi di dati non disgiunti. Perciò se in un intervallo di dati l'ipotesi nulla viene rifiutata a causa di un falso allarme, è assai probabile che ciò avvenga anche nell'intervallo successivo se questo ha una larga sovrapposizione col precedente. D'altra parte si è visto in precedenza come una larga sovrapposizione tra gli intervalli sia indispensabile per non dar luogo a cospicue perdite di potenza del test sull'intera serie storica. Il risultato è una forte correlazione tra le statistiche test e, conseguentemente, una tendenza dei falsi allarmi ad essere concentrati nel tempo.

In questa situazione il tasso di falsi allarmi κ_{fa} , ricavato dal valore atteso di $N_{H_1|H_0, n_t}$, pur rimanendo valido sul lungo periodo, non è del tutto adeguato per caratterizzare il comportamento della serie di test di ipotesi che si stanno effettuando. Infatti se su 10000 intervalli di 80 dati ciascuno, spostati di un solo dato l'uno rispetto all'altro, solo i primi 10 presentano un falso allarme, allora il rate è di un falso allarme ogni mille intervalli, ma di fatto la sequenza di dati da salvare per un'analisi successiva è una sola, estesa su poco meno di un centinaio di dati.

Una descrizione più adeguata della situazione che ci si trova ad affron-

tare con la strategia di test adottata è fornita dalla probabilità $P_{\text{fa}}(n_t) = \Pr(N_{H_1|H_0, n_t} > 0)$ di avere almeno un falso allarme in n_t intervalli, al variare di n_t stesso. Ancora una volta l'idea di adottare $P_{\text{fa}}(n_t)$ deriva dal campo dell'inferenza statistica simultanea, come spiegato in [35]. Si fa notare che l'esigenza di una stima di $P_{\text{fa}}(n_t)$ non si presenta soltanto per i test wavelet proposti nella sezione 2.2: si tratta piuttosto di un'esigenza comune per tutti i test proposti in letteratura per la rivelazione di segnali transienti. Tuttavia la definizione di questo problema, a quanto ci risulta, viene messa in luce per la prima volta in questa tesi.

Per $\Pr(N_{H_1|H_0, n_t} > 0)$ non esistono formule generali come per $\mathcal{E}[N_{H_1|H_0, n_t}]$, ed il suo calcolo nel contesto che stiamo trattando è risultato piuttosto complesso per le statistiche wavelet introdotte in questo lavoro. Anche per le altre statistiche presentate nella sezione 2.1 il calcolo non appare triviale.

Il problema che si pone è del tutto analogo a quello che si presenta per la *Scan Statistic*, descritta ad esempio in [47]. La *Scan Statistic* misura il numero degli eventi a cui si è interessati su tutti i possibili intervalli temporali di una prefissata larghezza n che sono contenuti entro un determinato lasso di tempo; lo scopo è quello di individuare intervalli con una concentrazione anomala di eventi in relazione al numero totale di eventi osservati o al tasso di eventi, supposto noto. L'analogia nasce dal fatto che anche nel caso della ricerca di segnali transienti si ricercano valori elevati assunti da una certa statistica su una successione di intervalli che copre un dato arco di tempo. Ciò che cambia tra i due casi è la statistica calcolata nei singoli intervalli e la sua distribuzione. D'altra parte la distribuzione cambia anche per la statistica di conteggio degli eventi usata nella *Scan Statistic*: binomiale nel caso in cui sia fissato il numero totale degli eventi, poissoniana nel caso in cui sia il tasso di eventi per unità di tempo ad essere fissato.

La linea di ragionamento sviluppata in [48] per la *Scan Statistic*, può essere utilizzata anche per cercare di stimare la probabilità $P_{\text{fa}}(n_t)$ definita in precedenza. Sia \mathcal{S} l'insieme degli n_t intervalli considerati e supponiamo che sia $\mathcal{C} \subset \mathcal{S}$ un sottoinsieme di n_c intervalli ad intersezione nulla e contigui tra loro, in grado di contenere tutti i dati contenuti negli intervalli di \mathcal{S} . Consideriamo

gli eventi:

$$\begin{aligned}
A_i &= \{T(\mathcal{I}_i^{(c)}) \geq \tau, \mathcal{I}_i^{(c)} \in \mathcal{C}\}, \\
D_i &= \{\exists \mathcal{I}_i^{(\phi)} \in (\mathcal{S} - \mathcal{C}) : T(\mathcal{I}_i^{(\phi)}) \geq \tau, \mathcal{I}_i^{(\phi)} \cap \mathcal{I}_i^{(c)} \neq \emptyset \neq \mathcal{I}_i^{(\phi)} \cap \mathcal{I}_{i+1}^{(c)}\}, \\
B_i &= D_i \cap \{\forall \mathcal{I}_i^{(c)} \in \mathcal{C}, T(\mathcal{I}_i^{(c)}) < \tau\},
\end{aligned} \tag{2.47}$$

dove τ è la soglia al di sopra della quale la statistica $T(\mathcal{I})$, valutata nell'intervallo \mathcal{I} , suggerisce di rifiutare l'ipotesi nulla H_0 . In sostanza A_i si verifica quando si accetta l'ipotesi alternativa H_1 nell' i -esimo degli n_c intervalli di \mathcal{C} , mentre B_i avviene quando H_1 viene accettata in un intervallo sovrapposto all' i -esimo e all' $(i + 1)$ -esimo intervallo di \mathcal{C} e contemporaneamente rifiutata in tutti gli intervalli di \mathcal{C} . Si verifica facilmente che, se vale l'ipotesi nulla, la probabilità che la soglia venga superata in almeno uno degli n_t intervalli di \mathcal{S} è data da:

$$P_{\text{fa}}(n_t) = \Pr(A) + \Pr(B), \tag{2.48}$$

dove si è posto $A = \cup_i A_i$ e $B = \cup_i B_i$. Dal momento che gli intervalli di \mathcal{C} sono disgiunti, sotto ipotesi nulla è facile mostrare che gli eventi A_i sono tra loro indipendenti e dunque $\Pr(A) = n_c \alpha$. In [48], applicando i limiti di Bonferroni modificati come in [50], si ottiene:

$$\begin{aligned}
\Pr(B) &\geq (n_c - 1)\Pr(B_1) - (n_c - 2)\Pr(B_1 B_2) - \frac{(n_c - 2)(n_c - 3)}{2}\Pr(B_1 B_3), \\
\Pr(B) &\leq (n_c - 1)\Pr(B_1) - (n_c - 2)\Pr(B_1 B_2).
\end{aligned} \tag{2.49}$$

Il calcolo delle probabilità $\Pr(B_1)$, $\Pr(B_1 B_2)$ e $\Pr(B_1 B_3)$ risulta in ogni caso di notevole difficoltà. Nel caso della Scan Statistic queste quantità vengono ricavate sfruttando pesantemente la natura discreta di questa statistica. Rimane invece un problema aperto la ricerca di un metodo che consenta di calcolare $\Pr(B_1)$, $\Pr(B_1 B_2)$ e $\Pr(B_1 B_3)$ per le statistiche discusse in questa tesi.

2.2.3 Considerazioni pratiche

L'implementazione delle statistiche descritte nella sezione 2.2 passa dalla capacità di calcolare i coefficienti wavelet $d_{j,m}^{(Y)}$ e quelli di scala $c_{j,m}^{(Y)}$ a partire dai dati della serie Y_k , con $k \in \mathbb{Z}$. D'altra parte non è realizzabile in pratica la trasformazione di infiniti dati in infiniti coefficienti; perciò qualsiasi algoritmo per il calcolo di trasformate wavelet, come ad esempio quello descritto in [34],

utilizza soltanto un insieme finito di dati contigui $\mathcal{Y} = \{Y_k : 1 \leq k \leq n\}$ per calcolare i coefficienti corrispondenti all'intervallo temporale dato da $1 \leq k \leq n$. Questa approssimazione risulta peggiore per i coefficienti più vicini ai margini dell'intervallo considerato: si parla pertanto di *effetto di bordo*. Infatti per i coefficienti ben all'interno dell'intervallo l'approssimazione può essere estremamente accurata o addirittura esatta nel caso in cui si usino wavelet madri e padri a supporto compatto, come quelle descritte in [15].

Un altro vantaggio di wavelet madri e padri a supporto compatto è la disponibilità di un algoritmo veloce per il calcolo dei coefficienti: nel caso in cui sia $n = 2^l$ si deve effettuare un numero di operazioni dell'ordine di $O(2^l)$ invece che dell'ordine di $O(l 2^l)$ come richiederebbe l'uso di wavelet a supporto non compatto (si veda [34], [21]). Se poi si usano valori di n diversi da potenze di 2 il costo computazionale cresce ancora; ecco perché si è sempre posto $n = 2^l$ anche nelle precedenti sezioni.

È bene osservare che le considerazioni sui costi computazionali non sono di scarso rilievo nella presentazione di una metodologia statistica. Infatti l'esigenza di analizzare considerevoli moli di dati in una prefissata unità di tempo, rende determinante la rapidità di calcolo per la praticabilità di una data metodologia nel contesto del problema da risolvere.

Nel caso dei dati prodotti dall'esperimento VIRGO l'analisi deve procedere di pari passo con l'acquisizione, che avverrà con frequenze di campionamento piuttosto elevate (10-20 KHz). Infatti, anche rinunciando ad una vera e propria analisi *on-line*, non si può permettere che il ritardo tra acquisizione ed analisi dei dati cresca indefinitamente al trascorrere del tempo. Perciò la ricerca della massima efficienza computazionale è fondamentale per una reale applicazione ai dati di VIRGO delle metodologie statistiche proposte nelle sezioni precedenti.

Alla luce delle considerazioni appena svolte, l'implementazione più naturale delle statistiche T descritte nella sezione 2.2 si ottiene calcolando la trasformata wavelet dei soli dati sotto esame Y_1, Y_2, \dots, Y_n ; il test viene quindi iterato spostando l'attenzione su intervalli di dati successivi. Per massimizzare la potenza complessiva del test gli intervalli consecutivi devono essere traslati di un solo dato alla volta. In effetti è questa l'implementazione che viene adottata ad esempio in [8].

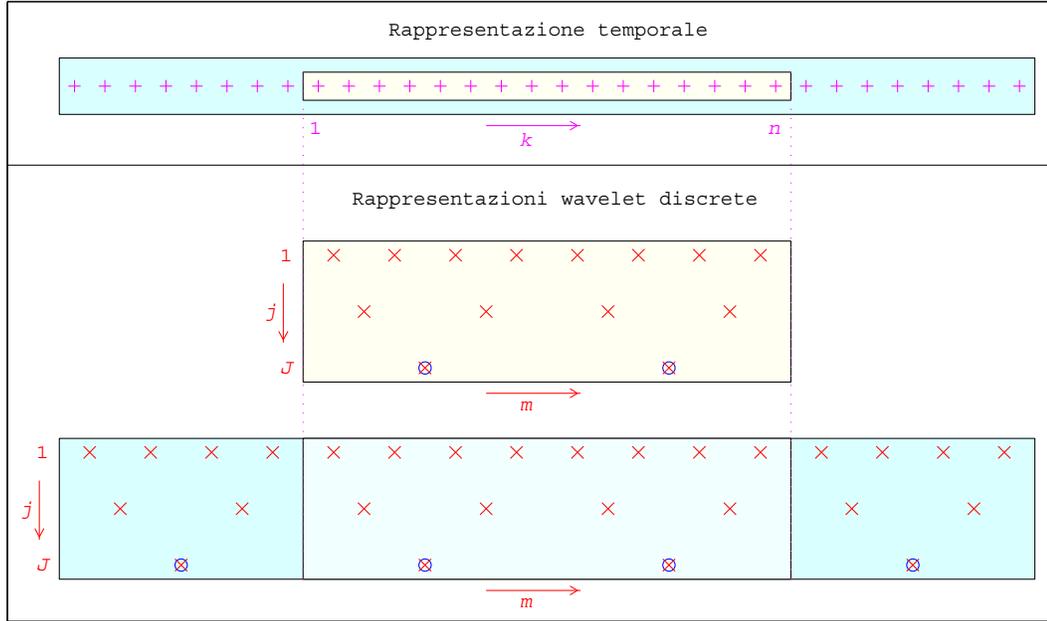


Figura 2.1: Disposizione temporale di una serie di dati e dei relativi coefficienti di una sua trasformata wavelet discreta. I coefficienti corrispondenti ai dati nell'intervallo centrale possono essere ottenuti sia dalla trasformata di tale intervallo (riquadro centrale) che dalla trasformata di un intervallo più esteso opportunamente dislocato (riquadro in basso).

In questa tesi invece si propone di adottare un'originale alternativa, partendo dall'osservazione che i coefficienti alla scala j di una trasformata wavelet discreta risultano covarianti per traslazioni di 2^j dati, benché non lo siano per traslazioni generiche. Perciò tutti i coefficienti fino alla scala J della trasformata di $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_n\}$ si ottengono, a meno di effetti di bordo, come sottoinsieme dei coefficienti della trasformata di un intervallo di dati molto più esteso \mathcal{Y}_e , purché tale intervallo inizi $h2^J$ dati prima di \mathcal{Y} , con $h \in \mathbb{N}$. In altre parole invece di isolare nel dominio temporale i dati che appartengono a \mathcal{Y} , si selezionano i coefficienti della rappresentazione wavelet di \mathcal{Y}_e che discendono dai dati in \mathcal{Y} . Questa situazione è raffigurata in figura 2.1.

A prima vista questa implementazione può mostrare soltanto l'aspetto negativo di richiedere un tempo di calcolo più elevato per calcolare i coefficienti relativi a \mathcal{Y} : infatti si è detto che il costo computazionale della trasformata wavelet discreta cresce con la lunghezza della serie dei dati analizzata. Tuttavia

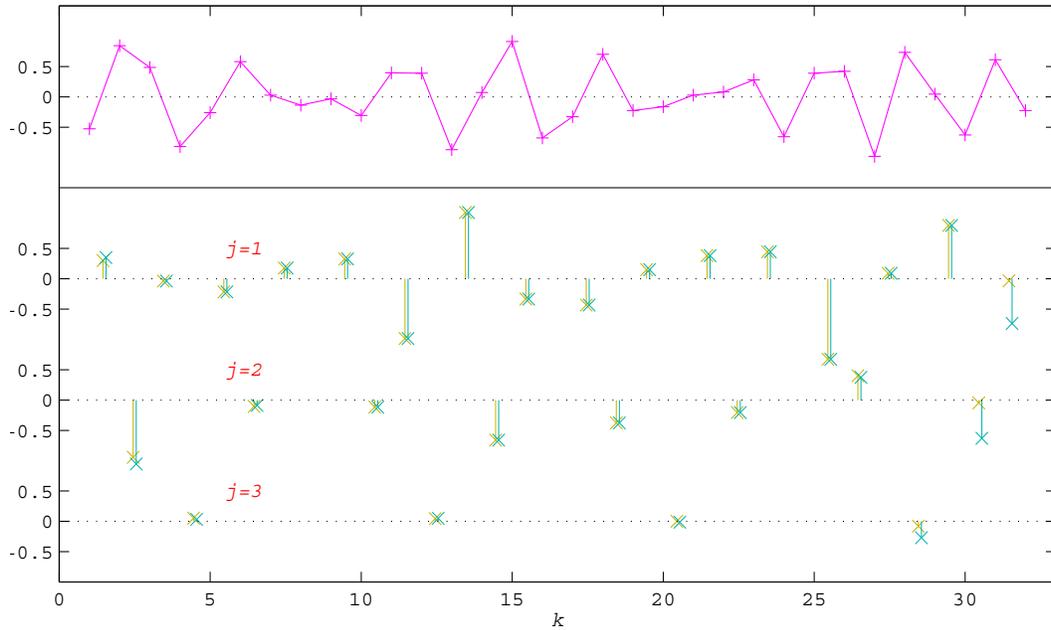


Figura 2.2: Serie di dati, in alto, e serie dei coefficienti wavelet corrispondenti per le prime tre scale. I coefficienti sono stati calcolati con entrambe le configurazioni presentate in figura 2.1: è evidente la discrepanza ai margini dell'intervallo tra i coefficienti calcolati usando solo i dati dell'intervallo, in giallo, e quelli ottenuti dalla trasformata di un intervallo più esteso, in verde-azzurro.

si ottengono alcuni considerevoli vantaggi, soprattutto in termini di rapidità di calcolo complessiva. Infatti questa implementazione:

1. riduce i tempi di calcolo complessivi per i test di ipotesi consecutivi su intervalli di analisi sovrapposti. Calcolata una volta per tutte la trasformata wavelet dei dati nell'intervallo esteso \mathcal{Y}_e , si ottengono i coefficienti fino alla scala J per *tutti* gli intervalli di n dati che, come \mathcal{Y} , sono contenuti in \mathcal{Y}_e ed hanno istante iniziale spostato di un multiplo di 2^J dati dall'istante iniziale di \mathcal{Y}_e . In pratica si sfrutta la citata proprietà di covarianza della trasformata wavelet discreta per traslazioni di multipli di 2^J dati, in modo da calcolare una sola volta i coefficienti che appartengono all'intersezione di intervalli sovrapposti. Invece, operando una trasformata wavelet discreta per ciascuno degli intervalli contenuti in \mathcal{Y}_e , i coefficienti delle intersezioni verrebbero ricalcolati una volta per ogni intervallo che li comprende.

2. consente di adottare gli algoritmi più veloci per il calcolo della trasformata wavelet discreta anche nel caso in cui la lunghezza degli intervalli analizzati non sia una potenza di 2, ma piuttosto un multiplo di 2^J . Infatti basta adottare una lunghezza che sia potenza di 2 per l'intervallo esteso \mathcal{Y}_e su cui viene effettivamente calcolata la trasformata wavelet. La lunghezza n dell'intervallo di analisi \mathcal{Y} costituisce uno dei parametri che caratterizzano le statistiche test proposte nella sezione 2.2: una maggiore flessibilità sui valori disponibili per la scelta di tale parametro può rappresentare un considerevole miglioramento per la potenza del test basato sulla statistica adottata. Per fare un esempio supponiamo di avere indicazioni a priori per un contributo rilevante, da parte del segnale transiente, della durata di circa 700-800 dati: adottando una lunghezza dell'intervallo di analisi di $2^9 = 512$ dati si rischia di perdere dei contributi importanti di segnale, mentre con una lunghezza di $2^{10} = 1024$ si può “diluire” troppo il segnale nel rumore. Ma l'implementazione con trasformata su \mathcal{Y}_e , a patto che la statistica test non usi coefficienti su scale $j > 8$, permette di adottare anche una lunghezza di $3 \cdot 2^8 = 768$ dati per l'intervallo di analisi, una scelta che può rappresentare un buon abbinamento con le caratteristiche del segnale transiente da ricercare.

3. consente di utilizzare soltanto coefficienti esenti da effetto di bordo. Infatti i coefficienti che risentono di tale effetto sono solo quelli vicini alle estremità dell'intervallo \mathcal{Y}_e e non quelli in prossimità dei bordi dell'intervallo di analisi \mathcal{Y} , purché \mathcal{Y} si trovi ben all'interno di \mathcal{Y}_e . Se invece avessimo calcolato i coefficienti relativi a \mathcal{Y} per mezzo della trasformata dei soli dati appartenenti a tale intervallo, allora l'effetto di bordo influenzerebbe in ogni caso i coefficienti in corrispondenza con i margini di \mathcal{Y} . In figura 2.2 si mostra la differenza tra i coefficienti calcolati con le due diverse implementazioni in una situazione analoga a quella mostrata in figura 2.1.

A fronte dei vantaggi appena elencati, l'implementazione con trasformata sull'intervallo esteso \mathcal{Y}_e presenta lo svantaggio di imporre intervalli, per test di ipotesi consecutivi, traslati l'uno rispetto all'altro di 2^J dati, con J dato dalla scala più grande utilizzata nella statistica test adottata. D'altra parte ciò implica una diminuzione della potenza del test, perchè non è detto che

il test venga effettuato sull'intervallo di n dati che garantirebbe la massima potenza: per essere sicuri di effettuare il test anche su tale intervallo occorrerebbe traslare la “finestra” di analisi di un dato alla volta. D'altra parte, se J rimane piccolo, soprattutto rispetto a $l = \log_2(n)$, ci si può aspettare che non si verifichino cospicue perdite di potenza se si ha una separazione di 2^J dati tra gli intervalli analizzati in successione. Altrimenti si può ricorrere a più intervalli estesi come \mathcal{Y}_e che coprano approssimativamente lo stesso intervallo di dati, ma leggermente traslati l'uno rispetto all'altro; tuttavia questa strategia comporta una moltiplicazione del costo di calcolo per il numero degli intervalli estesi utilizzati.

2.3 Test in presenza di rumore correlato

La serie di rumore X del modello 2.1 è stata considerata fino a questo punto come non correlata, gaussiana e localmente stazionaria. Tra queste assunzioni la prima appare come la più restrittiva per le applicazioni pratiche. È ragionevole ipotizzare che le condizioni di gaussianità e locale stazionarietà siano soddisfatte in un grande numero di situazioni, in particolare se il rumore è originato da una molteplicità di fenomeni fisici indipendenti con caratteristiche lentamente variabili nel tempo. Invece sono piuttosto rare le situazioni in cui il rumore si presenta senza alcuna struttura di correlazione. Si è dunque cercato di sviluppare una metodologia per l'applicazione delle statistiche della sezione 2.2 al caso di rumore correlato, mantenendo le altre assunzioni fatte.

Infatti restano valide le considerazioni qualitative alla base della definizione di queste statistiche test: anche in presenza di correlazione nella serie di rumore, sembra del tutto ragionevole cercare di rivelare un segnale transiente isolando, nella serie dei dati, l'intervallo temporale e le scale che lo contraddistinguono. Tuttavia, rispetto al caso di rumore bianco, non sono più applicabili i risultati della proposizione 2.5. D'altra parte, se non è nota la distribuzione della statistica sotto ipotesi nulla, allora è impossibile individuare, per un fissato valore dell'ampiezza α , la regione critica del test di ipotesi corrispondente.

È pertanto necessario formulare una proposizione analoga alla 2.5 che tenga conto della specifica struttura di correlazione che caratterizza il rumore

presente nella serie di dati da analizzare.

Il raggiungimento di questo obiettivo è reso assai complicato dalla natura frazionaria delle statistiche della sezione 2.2. Se si assume di conoscere esattamente la matrice di varianza e covarianza del rumore i denominatori di tali statistiche non sono di alcuna utilità: il loro ruolo era infatti quello di normalizzare le statistiche rispetto a fluttuazioni temporali della varianza σ_X^2 del rumore. Nel caso di rumore correlato avrebbe senso continuare ad usarle soltanto se la matrice di varianza e covarianza del rumore fosse nota a meno di una costante moltiplicativa.

Se ci si limita ad utilizzare solo i numeratori delle statistiche wavelet della sezione 2.2, il calcolo delle distribuzioni nel caso di rumore correlato dovrebbe risultare assai più semplice, anche se non triviale. Si tratta infatti di ricavare la distribuzione di una somma di variabili chi-quadro correlate tra loro.

L'approccio appena descritto cerca evidentemente di adattare i risultati della proposizione 2.5 alla struttura di correlazione dei dati. Viceversa si può pensare di trasformare i dati in modo che la loro struttura di correlazione risponda ai requisiti posti dalla proposizione 2.5. In questo caso la serie di dati originaria viene sottoposta ad un algoritmo di *whitening* che, in assenza di segnale, la trasformi in una serie di rumore bianco. In presenza di un segnale transiente invece anche la serie ottenuta con il *whitening* rimane correlata per un certo lasso di tempo. Ed è proprio sulla serie di dati prodotta dal *whitening* che si può continuare ad utilizzare i test delle sezioni 2.2 e 2.1. Naturalmente i segnali da rivelare si manifesteranno nella serie “sbiancata” con un aspetto diverso rispetto a quello originario; tuttavia si è in grado di determinare le caratteristiche di questi segnali trasformati a partire da quelle dei segnali originari.

In [1] vengono descritti alcuni algoritmi di *whitening* per una serie di rumore descritta, nella sua componente stazionaria, da un modello autoregressivo o ARMA: conoscendo i parametri di tale modello si risale alla serie di incrementi casuali ϵ_k che hanno originato la serie di dati osservati. In assenza di segnale questa serie di incrementi deve avere le caratteristiche di un rumore bianco e pertanto su questa serie possono essere implementati i test delle sezioni 2.1 e 2.2 per discernere tra le due ipotesi del modello 2.1.

Un altro metodo, estremamente intuitivo, per trasformare una serie di ru-

more correlato in una serie di rumore bianco passa dalla conoscenza della densità di potenza spettrale $S_f^{(X)}$ della serie di rumore X_k . Si propone la trasformazione:

$$\check{Y}_k = \frac{1}{2\pi} \sum_{\{f: S_f^{(X)} > 0\}} \frac{Y_f}{\sqrt{S_f^{(X)}}} \exp[i2\pi f k], \quad (2.50)$$

dove Y_f è la trasformata di Fourier discreta della serie originaria di dati Y_k . Data la linearità della trasformata di Fourier il modello 2.1 può essere riscritto come:

$$\check{Y}_k = \begin{cases} \check{X}_k, & \text{se il segnale non è presente (ipotesi nulla } H_0); \\ \check{X}_k + \check{s}_k, & \text{altrimenti (ipotesi alternativa } H_1), \end{cases} \quad (2.51)$$

dove \check{X}_k e \check{s}_k sono ottenute applicando la trasformazione 2.50 alle serie X_k ed s_k rispettivamente. Si dimostra facilmente che la serie \check{X}_k ha le caratteristiche di un rumore bianco. Infatti la sua densità di potenza spettrale è data da:

$$S_f^{(\check{X})} = |\check{X}_f|^2 = \frac{|X_f|^2}{S_f^{(X)}} = \frac{S_f^{(X)}}{S_f^{(X)}} = 1, \quad (2.52)$$

dove \check{X}_f e X_f sono le trasformate di Fourier discrete delle serie \check{X}_k e X_k rispettivamente. Perciò i test di ipotesi delle sezioni 2.1 e 2.2 possono essere applicati senza alcuna modifica al modello 2.51 anziché direttamente al modello equivalente 2.1.

In realtà tutti gli approcci descritti per il caso di rumore correlato nascondono tre problemi. In primo luogo si assume di conoscere esattamente le struttura di correlazione del rumore, richiedendo di volta in volta o la matrice di varianza e covarianza, o i parametri di un modello autoregressivo/ARMA, o la densità di potenza spettrale del rumore. Nella pratica reale è assai improbabile che queste grandezze siano note a priori e dunque occorre stimarle dai dati stessi. D'altra parte i risultati enunciati in precedenza valgono rigorosamente solo se si utilizzano le quantità vere mentre debbono essere considerati approssimativi qualora si faccia ricorso a quantità stimate.

Il secondo problema è strettamente connesso con la stima delle proprietà di correlazione del rumore. Infatti per effettuare questa stima occorrerebbe disporre di sequenze di rumore puro, senza alcun segnale sovrapposto. Se gli unici dati disponibili sono quelli della serie sotto esame, l'assenza di segnali

non può essere garantita a priori, altrimenti non ci sarebbe bisogno di alcun test di ipotesi per rigettare l'ipotesi alternativa dei modelli 2.1 e 2.51.

A questo tipo di inconveniente si può cercare di porre rimedio utilizzando metodologie di stima dei parametri del rumore che siano robuste rispetto all'eventuale presenza di segnali sovrapposti di breve durata. Ad esempio la densità di potenza spettrale $S_f^{(X)}$, richiesta dalla trasformazione 2.50, può essere stimata con una mediana delle stime in più intervalli temporali diversi da quello sotto analisi: in tal modo l'effetto di una stima distorta dalla presenza di un eventuale segnale transiente in un intervallo verrebbe diluito dalle stime in intervalli in cui il transiente non si manifesta. Si assume infatti che i segnali transienti si manifestino con una relativa rarità.

Infine si pone un problema qualora non si intenda rimpiazzare l'ipotesi di locale stazionarietà per la serie di rumore a favore di una più restrittiva ipotesi di stazionarietà in senso debole. Infatti le stime fatte per matrici di varianza e covarianza, parametri autoregressivi, spettri, sono valide soltanto entro l'arco di tempo in cui la serie di rumore può essere considerata come stazionaria. Con l'ipotesi di stazionarietà solo locale si ammette però che queste quantità siano lentamente variabili nel tempo e non possano dunque essere stimate una volta per tutte. Si deve procedere ad una nuova stima ad intervalli regolari di tempo, con una durata dettata dalla specifica formulazione del generico concetto di locale stazionarietà. Altrimenti si possono utilizzare stime adattive che si aggiornino gradualmente nel tempo sotto l'influenza degli ultimi dati acquisiti e con un peso sempre minore attribuito ai dati più vecchi. Un esempio di stima adattiva di coefficienti autoregressivi viene suggerito in [1] e sviluppato in [13] per l'implementazione nell'ambito del progetto VIRGO.

Capitolo 3

Rivelazione di segnali gravitazionali da supernovae

In questo capitolo ci si propone di verificare quale sia l'effettiva utilità applicativa dei test di ipotesi che utilizzano le statistiche introdotte nella sezione 2.2. In particolare verrà fatto riferimento al caso della rivelazione di una particolare classe di segnali da parte di un interferometro per onde gravitazionali come VIRGO, confrontando i risultati con quanto già disponibile in letteratura, ad esempio in [28].

3.1 Criteri di valutazione delle prestazioni

Le prestazioni di un test di ipotesi usualmente vengono espresse per mezzo della potenza π del test sotto ipotesi alternativa H_1 , per un fissato valore α dell'ampiezza del test. Tale valore $\pi = \Pr(H_1|H_1)$ esprime la probabilità che il test suggerisca il rifiuto dell'ipotesi nulla quando questa effettivamente è falsa: è chiaro che un test risulta tanto migliore quanto più $\pi \rightarrow 1$.

L'ipotesi alternativa H_1 nel modello 2.1 è composta dall'unione di tutte le ipotesi semplici che assumono la sovrapposizione di uno specifico segnale s alla serie storica del rumore. Indicando con \mathcal{S} l'insieme di questi possibili segnali s , a cui si riferisce complessivamente H_1 , si ha:

$$\pi = \sum_{s \in \mathcal{S}} \Pr(H_1|H_1, s) \Pr(s). \quad (3.1)$$

ammesso che l'insieme \mathcal{S} sia numerabile, altrimenti occorre una formulazione corrispondente nel continuo. Per le statistiche proposte nella sezione 2.2 que-

sta formulazione può essere espressa in modo più conveniente per mezzo del parametro $\delta^{(s)}$, definito dall'espressione 2.41; infatti questo è l'unico parametro attraverso il quale il segnale modifica la distribuzione della statistica test T rispetto all'ipotesi nulla:

$$\pi = \int_{\{\delta^{(s)}:s \in \mathcal{S}\}} \Pr(H_1|H_1, \delta^{(s)}) f_\delta(\delta^{(s)}) d\delta^{(s)} \quad (3.2)$$

dove $f_\delta(\cdot)$ è la densità di probabilità per il valore del parametro $\delta^{(s)}$.

Il valore di $\Pr(H_1|H_1, \delta^{(s)})$, sulla base di quanto enunciato dalla proposizione 2.5, può essere ricavato da una distribuzione di Fisher non centrale. L'insieme $\{\delta^{(s)} : s \in \mathcal{S}\}$ è assai più facile da specificare di quanto non sia \mathcal{S} , essendo tipicamente costituito da un intervallo di valori in \mathbb{R}^+ ammissibili per $\delta^{(s)}$, nel caso più generale $0 < \delta^{(s)} < +\infty$. Tuttavia di norma sarà praticamente impossibile riuscire a specificare la distribuzione $f_\delta(\cdot)$, soprattutto in considerazione del fatto che le statistiche della sezione 2.2 sono state introdotte per rivelare segnali non ben conosciuti, appartenenti ad una famiglia \mathcal{S} assai estesa e non ben definita.

Il problema della definizione troppo vaga dell'insieme \mathcal{S} è stato aggirato, nel contesto del progetto VIRGO, introducendo un sottoinsieme standard di segnali $\mathcal{S}^* \subset \mathcal{S}$ su cui valutare e confrontare le prestazioni dei vari test di ipotesi descritti nella sezione 2.1. Si vedano a tale proposito i lavori [4], [5], [6], [9], [28]. L'insieme \mathcal{S}^* scelto risulta costituito da 78 segnali diversi, frutto dei modelli teorici proposti in [52] per l'emissione di onde gravitazionali prodotte in un'esplosione di supernova¹ a 1 Kpc di distanza dalla terra², con collasso del nucleo centrale a stella di neutroni. Tale insieme costituisce dunque un banco di prova particolarmente interessante, dal momento che la rivelazione del segnale gravitazionale generato in eventi di questo tipo è uno degli obiettivi principali del progetto VIRGO.

Si deve tuttavia considerare che i modelli teorici che portano alla definizione dei segnali in \mathcal{S}^* non sono particolarmente affidabili e sono ben lontani

¹Una supernova è la manifestazione dell'evento esplosivo con cui si conclude la vita di stelle sufficientemente massicce. Normalmente il residuo centrale lasciato dall'esplosione è un oggetto estremamente compatto, detto stella di neutroni, ma nel caso in cui la stella progenitrice sia di massa molto elevata può rimanere soltanto un buco nero, ed in alcuni casi forse niente.

²1 Kpc = 1 Kiloparsec = $3.084 \cdot 10^{16}$ Km

dall'essere gli unici disponibili. Inoltre esistono esplosioni di supernovae che avvengono con modalità e parametri molto diversi da quelli considerati in [52]. In effetti la fisica teorica allo stato attuale non è in grado di produrre un modello soddisfacente ed unanimemente accettato per l'emissione di onde gravitazionali in eventi di questo tipo.

In ogni caso le statistiche descritte nelle sezioni 2.1 e 2.2 non sono state introdotte al solo scopo di rivelare un determinato segnale di origine astrofisica: ad esse è demandato il compito di rivelare qualsiasi non stazionarietà di breve durata che vada a sovrapporsi al rumore localmente stazionario dello strumento. L'identificazione della natura degli eventi rivelati come reali segnali gravitazionali o qualche tipo di "voce" strumentale è rimandata ad una seconda fase di analisi, possibile soltanto quando un buon numero di eventi transienti sarà già stato osservato, consentendo la realizzazione di un catalogo dei tipi di transienti più comuni. Si ricorda che per "voce" strumentale si intende qualsiasi manifestazione di non stazionarietà nella serie analizzata che sia originata da eventi transienti nel dispositivo sperimentale.

Le considerazioni appena svolte indicano che la difficoltà a fare inferenza nasce dalla mancanza di una corretta specificazione dello spazio campionario \mathcal{S} e della misura di probabilità ad esso associata. Tale difficoltà viene aggirata riducendo lo spazio campionario a $\mathcal{S}^* \subset \mathcal{S}$, ma questa operazione è chiaramente piuttosto arbitraria. Affinché si possa sperare che i risultati dell'inferenza su \mathcal{S}^* siano rappresentativi di una classe molto più estesa di segnali (eventi) in \mathcal{S} , occorre almeno che questi risultati siano tra loro omogenei. Ciò vuol dire che è necessario esprimere un giudizio sulla robustezza delle procedure impiegate, oltre che sulla loro potenza π , preferendo tra di esse quelle che risultano più robuste in \mathcal{S}^* .

Nella letteratura citata non si utilizza π come misura delle prestazioni dei test di ipotesi considerati. Si preferisce piuttosto definire un indice legato alla distanza massima a cui può essere posta la sorgente di un dato segnale $s \in \mathcal{S}^*$ senza che il test di ipotesi che viene studiato commetta un errore del secondo tipo, accettando erroneamente l'ipotesi nulla H_0 di solo rumore nei dati. Questo indice presenta il vantaggio di risultare estremamente intuitivo e significativo nel contesto della rivelazione di segnali di origine astrofisica.

Alcuni dei segnali in \mathcal{S}^* risulteranno intrinsecamente più deboli di altri,

e saranno pertanto osservabili soltanto a distanze inferiori, qualsiasi sia la metodologia impiegata per la rivelazione. Occorre pertanto definire un indice diverso dalla massima distanza di rivelazione, in modo da rendere confrontabili tra loro i risultati ottenuti per diversi segnali $s \in \mathcal{S}^*$. A tale scopo si usano come valori di normalizzazione le distanze massime di rivelazione ottenibili con il test del rapporto di verosimiglianza, per ogni specifico segnale $s \in \mathcal{S}^*$.

Nel contesto della rivelazione di segnali il test del rapporto di verosimiglianza viene spesso chiamato col nome di *matched filter* o filtro di Wiener. Nella sezione seguente daremo una descrizione dettagliata di questo test, che, come si vedrà, è utile come pietra di paragone ideale nel problema affrontato in questo lavoro, ma risulta non applicabile nel caso reale. Nella sezione successiva si approfondirà invece l'idea della distanza massima di rivelazione come indice delle prestazioni di una data statistica test.

3.1.1 Test del rapporto di verosimiglianza

Supponendo di conoscere esattamente tutte le caratteristiche del segnale $s = (s_1, s_2, \dots, s_n)$ da rivelare in una serie di dati $Y = (Y_1, Y_2, \dots, Y_n)$, entrambe le ipotesi del modello 2.1 sono ipotesi semplici. In questo caso il lemma di Neyman-Pearson indica il test del rapporto di verosimiglianza semplice come il test più potente tra tutti quelli con la stessa ampiezza α . Nel caso in cui il rumore sia gaussiano e con matrice di varianza e covarianza Σ nota, allora è in effetti possibile esprimere analiticamente il rapporto di verosimiglianza per il modello 2.1; si ha infatti:

$$\Lambda = \frac{\exp \left[-\frac{1}{2} Y' \Sigma^{-1} Y \right]}{\exp \left[-\frac{1}{2} (Y - s)' \Sigma^{-1} (Y - s) \right]} \quad (3.3)$$

che può essere riscritto come:

$$\Lambda = \exp \left[-\frac{1}{2} \left(s' \Sigma^{-1} Y + Y' \Sigma^{-1} s - s' \Sigma^{-1} s \right) \right]. \quad (3.4)$$

Dal momento che Σ^{-1} è simmetrica e definita positiva, per il teorema di Cholesky esiste una matrice triangolare Γ tale che $\Sigma^{-1} = \Gamma' \Gamma$. Ponendo $\tilde{s} = \Gamma s$ e $\tilde{Y} = \Gamma Y$, ricordando che il prodotto scalare è commutativo:

$$\Lambda = \exp \left[-\tilde{s}' \tilde{Y} + \frac{\|\tilde{s}\|^2}{2} \right], \quad (3.5)$$

e dunque la statistica:

$$W = \tilde{s}'\tilde{Y}, \quad (3.6)$$

detta filtro di Wiener o *matched filter*, è statistica sufficiente per Λ . Perciò esiste un valore di soglia $\tau_{\alpha,W}$ tale che il seguente test di ipotesi:

$$\begin{aligned} &\text{si accetti } H_0 \text{ qualora sia } W < \tau_{\alpha,W}, \\ &\text{si accetti } H_1 \text{ qualora sia } W \geq \tau_{\alpha,W}, \end{aligned}$$

ha ampiezza α e risulta essere quello di potenza massima tra tutti quelli che hanno la stessa ampiezza. Si veda in proposito [42].

Proposizione 3.1 *La statistica W è normalmente distribuita:*

$$W \sim \begin{cases} N_{0, \|\tilde{s}\|^2}, & \text{sotto } H_0; \\ N_{\|\tilde{s}\|^2, \|\tilde{s}\|^2}, & \text{sotto } H_1. \end{cases} \quad (3.7)$$

Dim: Sotto l'ipotesi nulla \tilde{Y} è distribuito come una normale multivariata di media nulla e matrice di varianza e covarianza data da:

$$\Gamma\Sigma\Gamma' = \Gamma\Gamma^{-1}(\Gamma^{-1})'\Gamma' = I. \quad (3.8)$$

Perciò la statistica $W = \tilde{s}'\tilde{Y}$, in quanto combinazione lineare di normali standard indipendenti, sotto H_0 ha distribuzione normale di media nulla e varianza pari a $\sum_{k=1}^n \tilde{s}_k^2 = \|\tilde{s}\|^2$. Sotto l'ipotesi alternativa H_1 si può scrivere $W = \tilde{s}'\tilde{X} + \|\tilde{s}\|^2$, dove per $\tilde{X} = \Gamma X$ valgono gli stessi ragionamenti svolti per \tilde{Y} nel caso dell'ipotesi nulla. In questo caso W risulta distribuita come una normale di media $\|\tilde{s}\|^2$ e varianza uguale a quella della distribuzione sotto H_0 . *qed*

Corollario 3.2 *Il valore di soglia $\tau_{\alpha,W}$ per il test del rapporto di verosimiglianza semplice di ampiezza α basato sulla statistica W è dato da:*

$$\tau_{\alpha,W} = \Phi_{(1-\alpha)}\|\tilde{s}\|, \quad (3.9)$$

dove $\Phi_{(1-\alpha)}$ è il quantile per la probabilità $(1-\alpha)$ di una distribuzione normale standard.

Il test del rapporto di verosimiglianza risulta ottimale soltanto nel caso in cui il segnale, presente quando è verificata l'ipotesi H_1 del modello 2.1, sia uno solo e perfettamente conosciuto in tutte le sue caratteristiche, compresa l'esatta collocazione temporale. In questo senso gli n dati utilizzati nel filtro di Wiener sono tutti e soli quelli per cui il segnale è diverso da zero, individuabili a priori proprio grazie alla conoscenza dell'eventuale segnale da rivelare. Se invece i segnali da rivelare sono molti e non noti, come nel problema affrontato in questo lavoro, il test del rapporto di verosimiglianza non è applicabile.

Sebbene non sia applicabile per la rivelazione di un generico segnale di \mathcal{S}^* , il test del rapporto di verosimiglianza fornisce un limite superiore per la distanza a cui si può sperare di rivelare ciascuno specifico segnale $s \in \mathcal{S}^*$. Infatti tale test è quello di maggiore potenza tra tutti quelli che sfruttano l'esatta conoscenza del segnale da rivelare e pertanto, a maggior ragione (si veda [23]), garantirà prestazioni migliori rispetto a test che sfruttano una conoscenza assai più ridotta sui segnali da rivelare, come è il caso dei test del capitolo 2.

Nel caso di rumore correlato il filtro di Wiener richiede la conoscenza della matrice di varianza e covarianza del rumore Σ : ciò pone per questo test di ipotesi lo stesso tipo di problemi descritti nella sezione 2.3 per le statistiche delle sezioni 2.1 e 2.2.

Per completezza si deve infine ricordare che, nel caso in cui non si disponga di alcuna informazione sulle caratteristiche del segnale da rivelare, il test che garantisce le migliori prestazioni è il test del rapporto di verosimiglianza generalizzato, come indicato in [23]. In questo caso la statistica test è simile a W definita nell'equazione 3.6, ma con \tilde{s} sostituito dalla sua stima di massima verosimiglianza \tilde{Y} . Si ottiene così una statistica equivalente al *Norm Filter* descritto nella sezione 2.1, sebbene ciò non sia stato osservato dagli autori che hanno introdotto questa statistica in [4].

Tuttavia, nel caso preso in considerazione in questo lavoro, si ipotizza che almeno alcune vaghe conoscenze sui segnali da rivelare siano disponibili e perciò è lecito attendersi che esistano statistiche di maggiore potenza rispetto al *Norm Filter*. In effetti è questo il caso per i 78 segnali in \mathcal{S}^* , come evidenziato in [28] con un confronto tra i risultati delle statistiche test finora proposte in letteratura per lo specifico problema applicativo considerato.

3.1.2 Distanze soglia

I segnali forniti da [52] sono stati tutti calcolati ammettendo che la sorgente si trovasse ad una distanza di 10000 Kpc. Ad una distanza così elevata i segnali risultano estremamente deboli, e con le tecnologie attuali non sarebbero in alcun modo rivelabili. Essendo noto che l'ampiezza delle onde gravitazionali decresce linearmente con l'aumentare della distanza, nell'insieme \mathcal{S}^* i 78 segnali sono stati normalizzati per una distanza di 1 Kpc. Dunque se $z^{(i)}$, $i \in \{1, 2, \dots, 78\}$, è l' i -esimo segnale in \mathcal{S}^* , allora:

$$s_k^{(i,d)} = \frac{z_k^{(i)}}{d}, \quad (3.10)$$

è lo stesso segnale, emesso però da una sorgente a d Kpc dal punto di osservazione (il nostro pianeta).

Ci si può chiedere in effetti quale sia la massima distanza $d = D_{i,T}$ da usare in $s^{(i,d)}$ affinché il test di ipotesi fondato sulla statistica T accetti ancora l'ipotesi alternativa H_1 del modello 2.1. $D_{i,T}$ è chiaramente una variabile aleatoria, poiché il valore che assume dipende dalla particolare realizzazione del processo stocastico di rumore presente nella serie di dati analizzati. Pertanto come indice della capacità della statistica T di rivelare l' i -esimo segnale si può considerare la mediana $\hat{d}_{i,T}$ della distribuzione di $D_{i,T}$.

In realtà nei lavori [4], [5], [6], [9], [28], la variabile aleatoria $D_{i,T}$ non viene introdotta e di fatto ci si limita ad utilizzare il suo valore medio $\bar{d}_{i,T}$: ma questo indice è assai meno robusto della mediana rispetto a rivelazioni sulla coda della distribuzione di $D_{i,T}$ verso i valori più elevati e meno probabili. La differenza tra i due indici è comunque di poco rilievo per statistiche con distribuzione approssimativamente normale, anche se l'indice che utilizza la mediana dovrebbe risultare di maggiore affidabilità nel caso in cui la distribuzione di $D_{i,T}$ venga ricavata per mezzo di un numero non molto elevato di simulazioni.

Il valore di $\hat{d}_{i,T}$ consente un agevole confronto tra le prestazioni di statistiche diverse applicate alla rivelazione di uno stesso tipo di forma d'onda $z^{(i)} \in \mathcal{S}^*$. Tuttavia non sono possibili confronti tra le prestazioni ottenute dalla stessa statistica applicata a segnali di forma diversa. Infatti, come si è detto, alcuni segnali possono essere, a parità di distanza, intrinsecamente più deboli di altri e quindi per essi anche valori non elevati di $\hat{d}_{i,T}$ debbono essere considerati positivamente.

Per ottenere un indice che abbia lo stesso significato per tutti i segnali, si può pensare di confrontare il valore di $\hat{d}_{i,T}$, ottenuto per la statistica test T , con il valore $\hat{d}_{i,W}$ che si otterrebbe sullo stesso segnale utilizzando il test del rapporto di verosimiglianza. Si introduce quindi come indice di prestazioni della statistica T sulla forma d'onda $z^{(i)}$ il rapporto:

$$\varepsilon_{i,T} = \frac{\hat{d}_{i,T}}{\hat{d}_{i,W}}, \quad (3.11)$$

che assume valori nell'intervallo adimensionale $[0, 1]$, data l'ottimalità del test del rapporto di verosimiglianza.

3.2 Le prestazioni dei test wavelet

In [4], [5], [6], [9], [28], i test di ipotesi proposti vengono messi alla prova sui segnali in \mathcal{S}^* campionati a 20 KHz, assumendo un rumore bianco e gaussiano con varianza $\sigma_X^2 = 1.6 \cdot 10^{-41}$ e fissando un tasso medio di falsi allarmi per unità di tempo κ_{fa} pari a 72 falsi allarmi per ora, ossia uno ogni 50 secondi di dati acquisiti. Questi valori sono stati fissati in modo da riprodurre una configurazione sperimentale simile a quella che si prevede di utilizzare in VIRGO.

In realtà il rumore strumentale non sarà affatto bianco. Per giustificare la semplificazione adottata nella letteratura citata si fa osservare che nell'intervallo di frequenze $[0.2, 1]$ KHz lo spettro dovrebbe risultare piuttosto piatto e che i segnali gravitazionali originati in esplosioni di supernovae dovrebbero manifestarsi prevalentemente in questa banda di frequenze. Inoltre è previsto che la serie di dati venga sottoposta ad una procedura di *whitening* che elimini la correlazione nella serie di rumore. Il valore di σ_X^2 citato in precedenza è fissato approssimativamente in corrispondenza del minimo dello spettro di rumore previsto per lo strumento, che cade proprio nell'intervallo di frequenze indicato.

In questo lavoro si operano scelte conformi a quelle citate, in modo da mantenere un confronto diretto con i risultati riportati in letteratura.

Le scelte sui parametri delle statistiche test della sezione 2.2 sono invece state operate sulla base delle vaghe conoscenze a priori disponibili sulle caratteristiche generali dei segnali da rivelare, senza tenere conto degli specifici

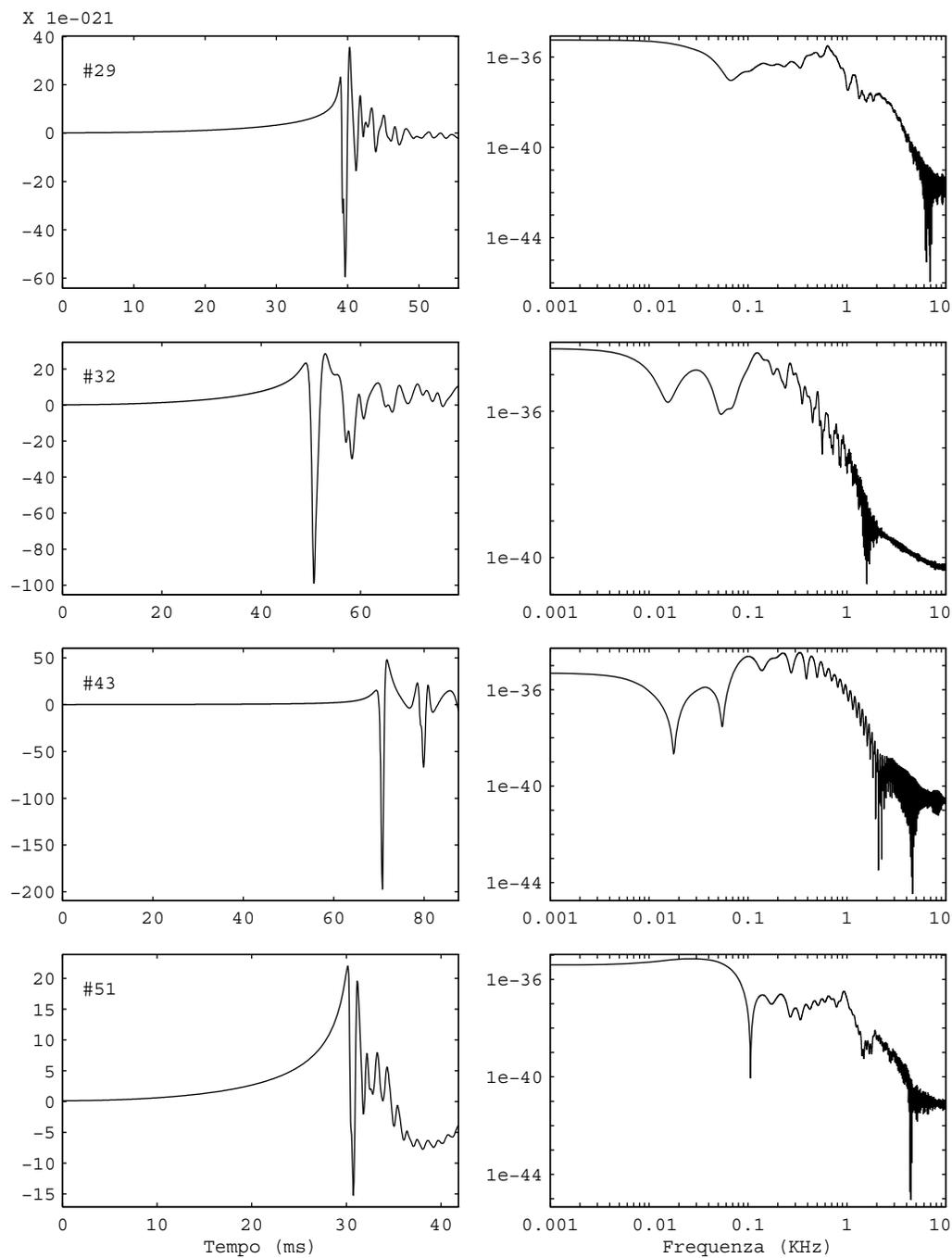


Figura 3.1: Alcuni dei 78 segnali in \mathcal{S}^* (a sinistra) assieme agli spettri di Fourier corrispondenti (a destra). Si noti che la base dei tempi e la scala verticale non è la stessa per tutti i segnali.

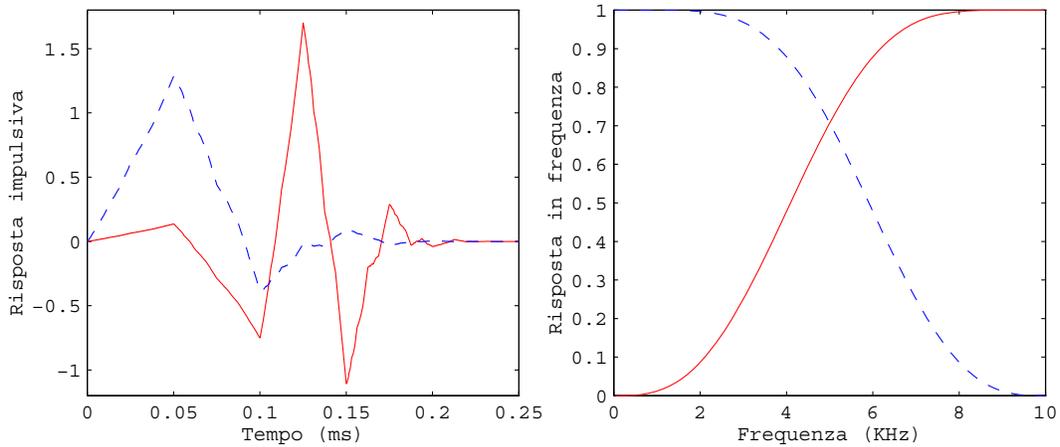


Figura 3.2: *A sinistra*: andamento temporale di una wavelet di Daubechies a minima asimmetria con 3 momenti nulli, linea continua, e della funzione di scala corrispondente, linea tratteggiata, per dati campionati ogni 0.05 millisecondi (20 KHz). *A destra*: risposte in frequenza delle due funzioni mostrate nella figura a sinistra.

78 segnali utilizzati in \mathcal{S}^* . Gli studi teorici sono in linea di massima concordi nell'attribuire ai segnali originati in esplosioni di supernovae un andamento oscillante con frequenze inferiori a pochi KHz, di durata variabile tra qualche millisecondo a qualche decina di millisecondi. In effetti queste indicazioni vengono rispettate anche dai segnali in \mathcal{S}^* , alcuni dei quali vengono mostrati, insieme agli spettri corrispondenti, in figura 3.1.

Si è scelto di usare come wavelet madre ortogonale ψ una wavelet di Daubechies ad asimmetria minima con i primi tre momenti nulli, descritta in [15] e mostrata in figura 3.2 assieme alla funzione di scala ϕ ad essa associata. Si tratta di una wavelet a supporto compatto e non molto esteso, con buone proprietà di localizzazione temporale dei picchi nel segnale. La risoluzione in frequenza non è molto elevata, come risulta dall'ampiezza della funzione di risposta mostrata sempre in figura 3.2; in ogni caso tale risoluzione è adeguata ai nostri scopi, dal momento che le oscillazioni dei segnali da rivelare sono in generale piuttosto irregolari, ben lontane da un andamento sinusoidale.

Per dati campionati a 20 KHz i coefficienti wavelet alla scala j colgono fluttuazioni nei dati a frequenze dell'ordine di $20/2^j$ KHz. Dal momento che si è ipotizzato per i segnali da rivelare un contributo rilevante soltanto per

frequenze inferiori a pochi KHz, si è scelto di utilizzare le prime due scale per i denominatori delle statistiche definite nella sezione 2.2. Infatti tali scale sono caratterizzate da frequenze di poco inferiori a 10 e 5 KHz rispettivamente, e quindi dovrebbero consentire un'adeguata stima della varianza del rumore senza essere influenzate dai segnali eventualmente presenti. Nei numeratori delle statistiche test vengono invece utilizzate combinazioni di coefficienti relativi a scale $j > 2$, corrispondenti a frequenze inferiori a 2.5 KHz, per le quali si presume che i segnali abbiano contributi non trascurabili.

Gli intervalli di dati analizzati in ogni singolo test di ipotesi contengono 256 dati ciascuno, ossia 12.8 millisecondi di dati campionati a 20 KHz. Questa larghezza risulta approssimativamente commisurata alle vaghe indicazioni sulle durate dei segnali da rivelare date in precedenza. Gli intervalli risultano separati l'uno dall'altro di 16 dati: ciò consente di sfruttare i benefici della trasformata su un intervallo esteso, descritti nella sezione 2.2.3, per calcolare statistiche test che utilizzino coefficienti di scale $j \leq 4$. Inoltre in questa configurazione la sovrapposizione tra intervalli consecutivi rimane notevole, 240 dati su 256: perciò si presume che sia limitata la perdita di potenza rispetto a test in intervalli spostati di un solo dato l'uno dall'altro.

In virtù delle considerazioni appena esposte si è scelto di verificare sui segnali in \mathcal{S}^* le prestazioni delle statistiche $T_{\psi_3,256}$, $T_{\psi_4,256}$, $T_{\phi_4,256}$, $T_{\psi_3\psi_4,256}$, $T_{\psi_4\phi_4,256}$, $T_{\psi_3\psi_4\phi_4,256}$. Con i valori di κ_{fa} e della frequenza di campionamento citati in precedenza, l'ampiezza dei singoli test di ipotesi, che si ottiene grazie all'equazione 2.46, risulta pari ad $\alpha = 1.6 \cdot 10^{-5}$: i valori di soglia che ne conseguono ed i gradi di libertà delle distribuzioni delle varie statistiche considerate sono riportati in tabella 3.1.

Per ognuna delle statistiche considerate e per ciascuno dei 78 segnali in \mathcal{S}^* è stata calcolata la probabilità di rivelazione $\Pr[T \geq \tau_{\alpha,T}|z^{(i)}, d]$ condizionata all'emissione della forma d'onda $z^{(i)} \in \mathcal{S}^*$ da parte di una sorgente posta ad una distanza di d Kpc. Si fa notare che questa probabilità in funzione di d non è altro che la distribuzione di probabilità cumulata da destra della distanza soglia $D_{i,T}$ introdotta nella sezione 3.1.2. La stessa quantità può essere interpretata anche come la potenza del test basato sulla statistica T per la forma di segnale $z^{(i)}$, al variare della distanza d della sorgente.

La probabilità $\Pr[T \geq \tau_{\alpha,T}|z^{(i)}, d]$ è stata ricavata sia teoricamente che per

	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$
ν_1	32	16	16	48	32	64
ν_2	192	192	192	192	192	192
$\tau_{\kappa_{\text{fa}}}$	0.4495	0.2923	0.2923	0.5969	0.4495	0.7397

Tabella 3.1: Gradi di libertà ν_1 e ν_2 e valori di soglia $\tau_{\kappa_{\text{fa}}}$ per le distribuzioni di Fisher che competono, sotto ipotesi nulla, ad alcune delle statistiche proposte nella sezione 2.2. Intervalli di 256 dati ciascuno separati di 16 dati l'uno dall'altro, $\kappa_{\text{fa}} = (1/50) \text{ sec}^{-1}$, ossia $\alpha = 1.6 \cdot 10^{-5}$. Wavelet e funzione di scala utilizzati sono mostrati in figura 3.2.

mezzo di simulazioni. Nel primo caso si è fatto ricorso alla distribuzione sotto ipotesi alternativa H_1 delle statistiche test fornita dalla proposizione 2.5. Il parametro di non centralità $\delta^{(s)}$ è stato ottenuto a partire da $z^{(i)}$ e d definendo il segnale come nell'espressione 3.10 ed applicando l'equazione 2.41. Nel secondo caso sono state utilizzate 1000 realizzazioni $x^{(r)}$ indipendenti e senza falsi allarmi della serie di rumore da sovrapporre a ciascun segnale, $r \in \{1, 2, \dots, 1000\}$. Esprimendo il segnale alla distanza d per mezzo della relazione 3.10 e ricordando la linearità della trasformata wavelet discreta enunciata nella proposizione 2.1, si ha:

$$|d_{j,m}^{(y^{(i,d,r)})}|^2 = \frac{|d_{j,m}^{(z^{(i)})}|^2}{d^2} + 2 \frac{d_{j,m}^{(z^{(i)})} d_{j,m}^{(x^{(r)})}}{d} + |d_{j,m}^{(x^{(r)})}|^2, \quad (3.12)$$

dove $y^{(i,d,r)}$ è la serie di dati per l' i -esimo segnale con sorgente a distanza d , sovrapposto all' r -esima realizzazione di rumore. Per mezzo della relazione 3.12 e dell'analogia per i coefficienti di scala, l'equazione $T = \tau_{\alpha,T}$, per le statistiche della sezione 2.2, si risolve come un'equazione di secondo grado in d : la radice positiva rappresenta il valore assunto dalla distanza soglia $D_{i,T}$ in corrispondenza della realizzazione $x^{(r)}$ della serie di rumore. In tal modo si ottiene una distribuzione empirica di 1000 valori della variabile aleatoria $D_{i,T}$.

In generale si trova un ottimo accordo tra i risultati delle simulazioni e le distribuzioni teoriche. Alcuni esempi vengono mostrati nella figura 3.3. In tabella 3.2 vengono riportate alcune statistiche riassuntive dei valori assunti, tra i 78 segnali di \mathcal{S}^* , dal rapporto:

$$\rho_{i,T} = \frac{\hat{d}_{i,T}^{(teor.)}}{\hat{d}_{i,T}^{(emp.)}} \quad (3.13)$$

tra la mediana teorica $\hat{d}_{i,T}^{(teor.)}$ e quella empirica $\hat{d}_{i,T}^{(emp.)}$ della distanza soglia $D_{i,T}$.

Sia nella tabella 3.2 che nella figura 3.3 è evidente, per la statistica $T_{\psi_3,256}$, una leggera sovrastima della mediana teorica rispetto a quella empirica. Questo effetto può essere meglio compreso andando a considerare il caso del segnale #16 in \mathcal{S}^* , per il quale $\rho_{16,T_{\psi_3,256}} \simeq 300$.

In figura 3.4 vengono mostrati, al variare dell'intervallo di analisi, i valori assunti dalle sommatorie dei moduli quadri dei coefficienti wavelet alle prime tre scale, sommatorie che compaiono nella definizione della statistica $T_{\psi_3,256}$ data dall'equazione 2.11. La serie di dati analizzata è costituita dal segnale #16 ad una distanza di 1 Kpc sovrapposto ad una particolare realizzazione, senza falsi allarmi, della serie di rumore; viene mostrato anche il valore assunto dalle sommatorie citate quando nei dati è presente soltanto il rumore o soltanto il segnale. È evidente che il contributo del segnale non è affatto trascurabile alla seconda scala, al contrario di quanto era stato assunto nella proposizione 2.5 per ricavare la distribuzione teorica della statistica test $T_{\psi_3,256}$. Considerando che la sommatoria relativa ai coefficienti della seconda scala veniva usata nella statistica test al fine di stimare la varianza del rumore, si capisce che la presenza del segnale comporta una notevole sovrastima di tale varianza. Questa sovrastima si produce proprio in corrispondenza di quegli intervalli in cui il segnale, contribuendo in modo significativo alla sommatoria relativa alla terza scala, dovrebbe portare la statistica test a superare il valore di soglia fissato per la rivelazione. Il risultato è quello di una totale cancellazione degli effetti

	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$
ρ_{\min}	0.9910	0.9854	0.9857	0.9841	0.9844	0.9833
$\bar{\rho}$	1.0307 [†]	0.9994	0.9984	1.0004	0.9970	0.9976
ρ_{\max}	1.2687 [†]	1.0215	1.0194	1.0284	1.0127	1.0122
σ_ρ	0.0522 [†]	0.0081	0.0068	0.0092	0.0060	0.0055

Tabella 3.2: Valori minimi, medi, massimi e deviazione standard dei 78 rapporti $\rho_{i,T}$, tra le mediane teoriche e quelle simulate della distribuzione della distanza soglia per il segnale # i rivelato dalla statistica T . Caratteristiche delle statistiche test come in tabella 3.1, varianza del rumore $\sigma_X^2 = 1.6 \cdot 10^{-41}$. Il simbolo † indica valori ottenuti con l'esclusione del segnale #16.

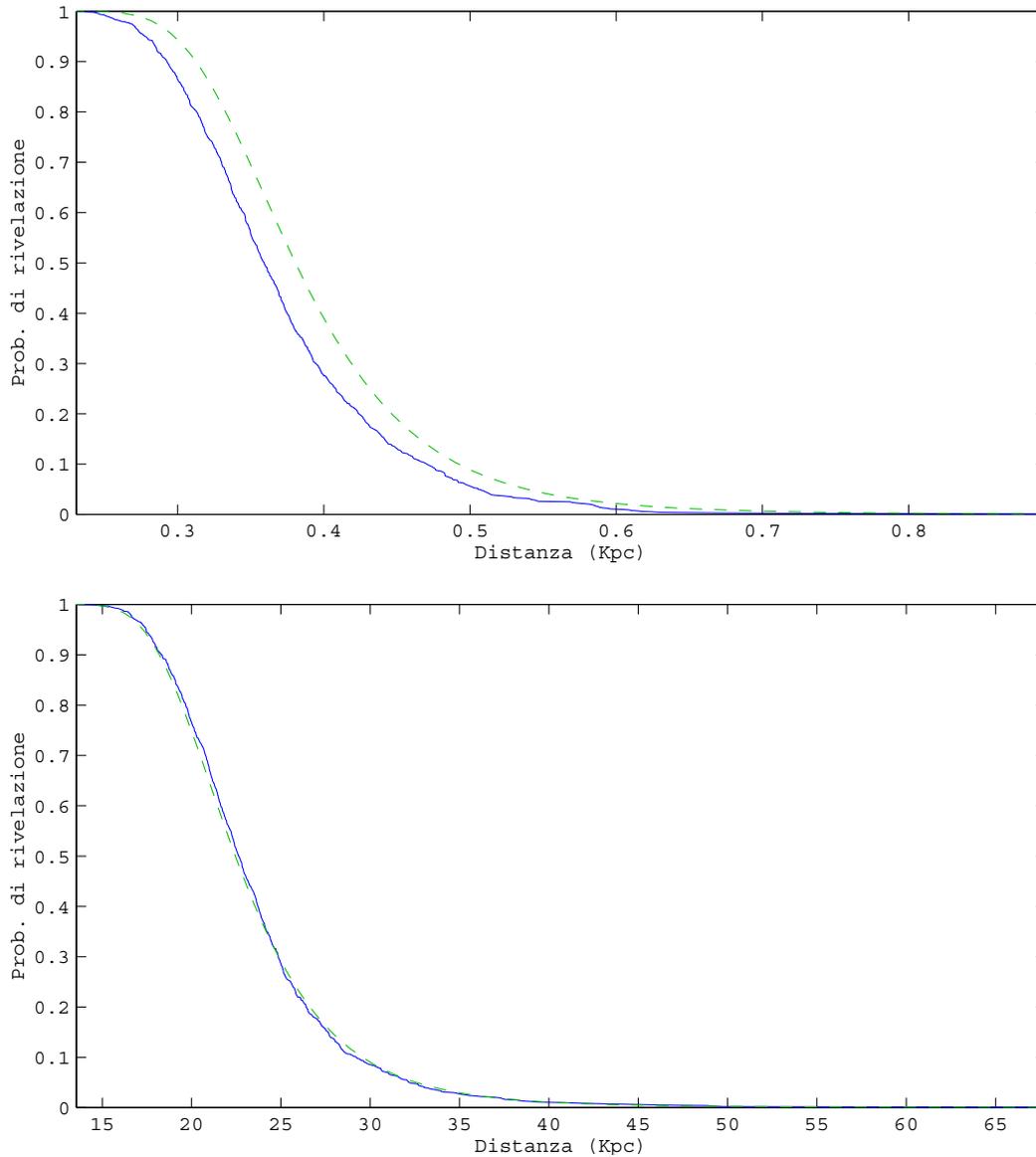


Figura 3.3: *In alto*: probabilità di rivelazione condizionata alla distanza della sorgente per il segnale #19 e la statistica $T_{\psi_3,256}$; la stessa quantità può essere interpretata come la distribuzione di probabilità, cumulata da destra, della distanza soglia $D_{19,T_{\psi_3,256}}$. La curva continua è il risultato di 1000 simulazioni di rumore senza falsi allarmi, la curva tratteggiata è ottenuta dalla distribuzione teorica. I parametri sono gli stessi impiegati per la tabella 3.2. *In basso*: come nella figura sopra, ma per il segnale #48 e la statistica $T_{\phi_4,256}$.

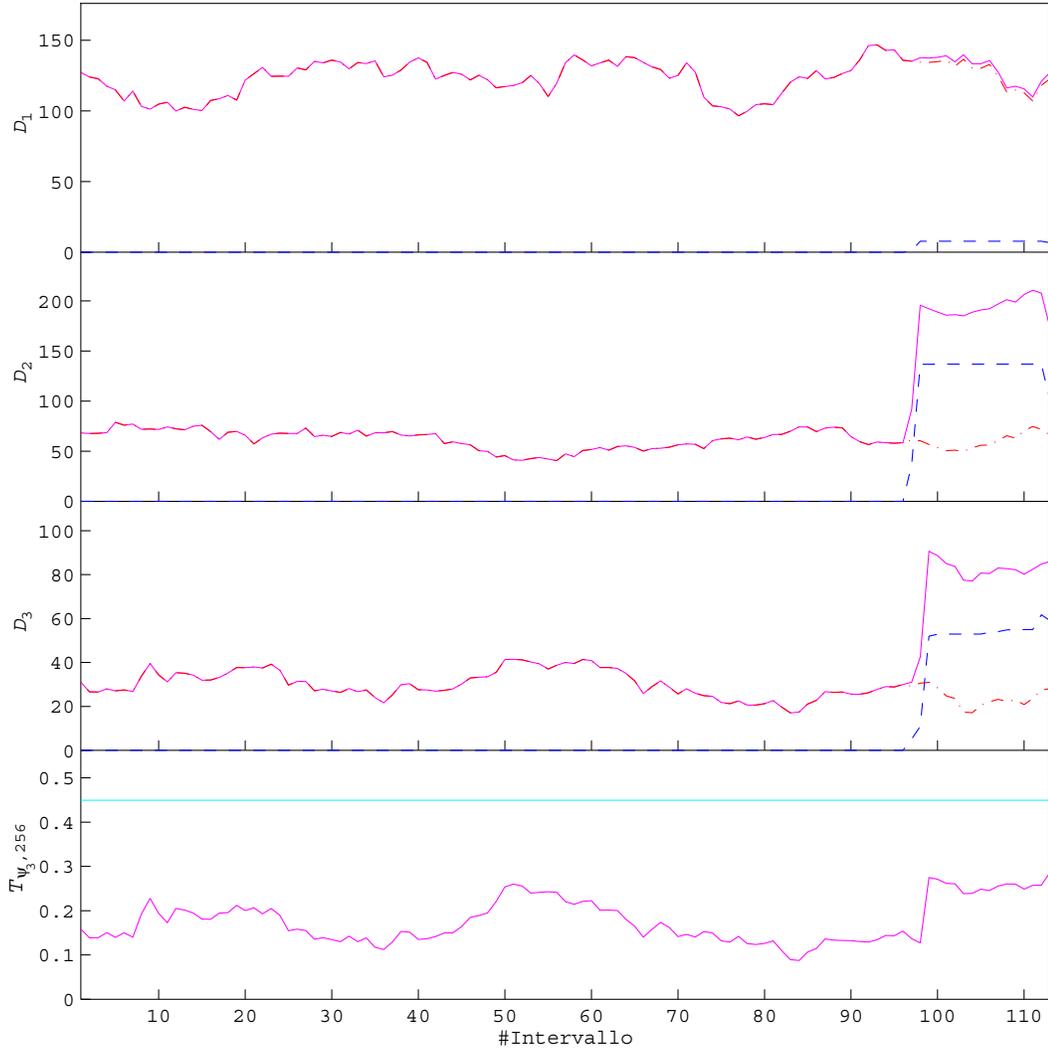


Figura 3.4: Nel riquadro superiore viene mostrato l'andamento della somma dei moduli quadri dei coefficienti wavelet alla scala $j = 1$ in intervalli di 256 dati ciascuno, separati di 16 dati l'uno dall'altro; nei due riquadri sottostanti la stessa quantità è mostrata per le scale $j = 2$ e $j = 3$. La linea continua si riferisce al segnale immerso nel rumore, mentre le linee tratteggiate e a tratto-punto sono, rispettivamente, per il solo segnale ed il solo rumore. Il segnale considerato è il #16 ad una distanza di 1 Kpc, la mediana della distribuzione teorica della distanza soglia per la statistica $T_{\psi_3,256}$; l'effettivo valore assunto nei vari intervalli da questa statistica viene confrontato nel quarto riquadro con il relativo valore soglia $\tau_{k_{fa}}$. La presenza del segnale è evidente alla seconda ed alla terza scala, ma non nella statistica test. I parametri utilizzati sono gli stessi indicati per la tabella 3.2.

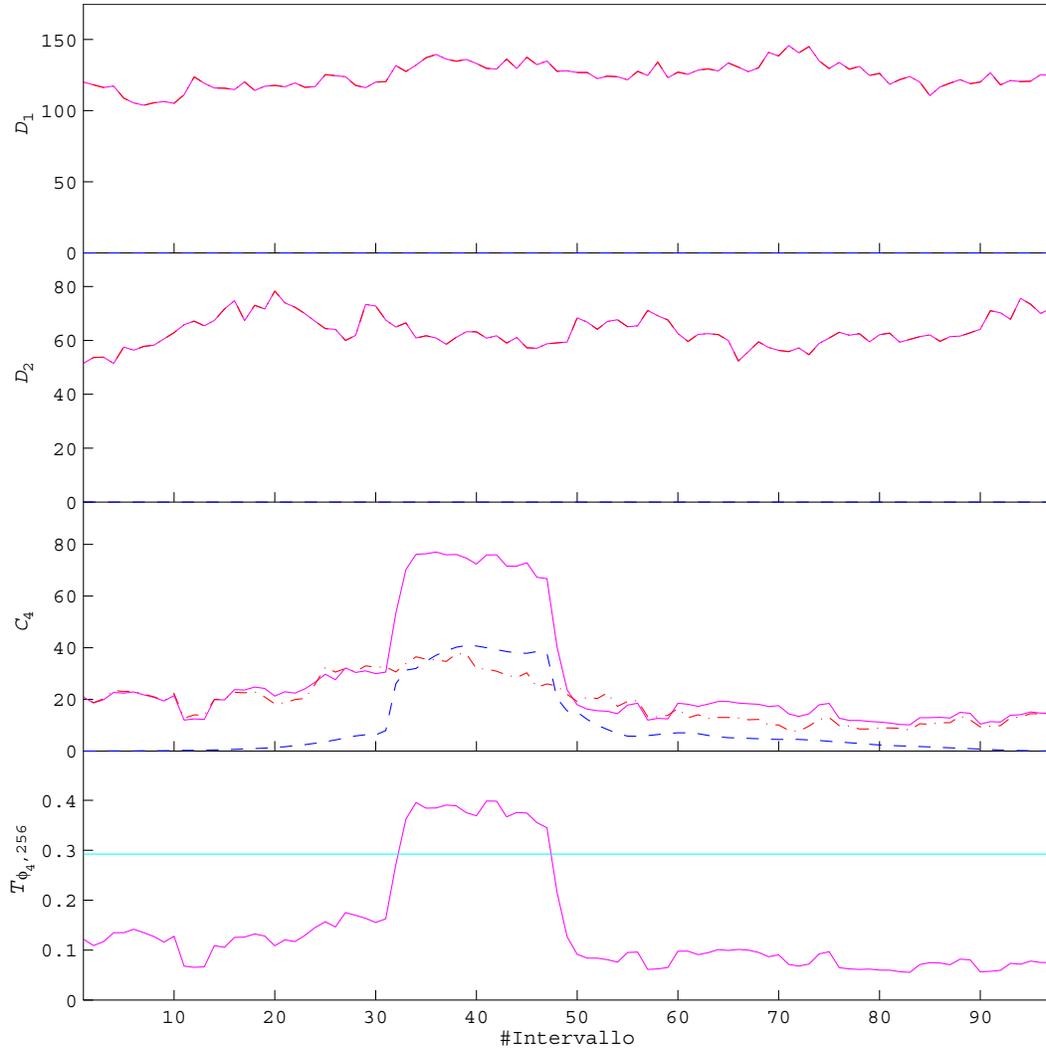


Figura 3.5: Grafici come in figura 3.4, ma per il segnale #75 ad una distanza di 35 Kpc e per la statistica $T_{\phi_4, 256}$; si noti che nel terzo riquadro è valutata la somma dei moduli quadri dei coefficienti di scala per $J = 4$, che compare come numeratore di $T_{\phi_4, 256}$.

	$T_{\psi_2,256}$	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_2\psi_3,256}$
ν_1	64	32	16	16	96
$\tau_{K_{fa}}$	1.1907	0.7148	0.4600	0.4600	1.6516
$\hat{d}_{16}^{(t)}$	1.4778	1.1372	12.9913	21.4441	1.5336
$\hat{d}_{16}^{(e)}$	1.4219	1.0767	13.0768	21.4347	1.4736
ρ_{16}	1.0393	1.0563	0.9935	1.0004	1.0407

	$T_{\psi_3\psi_4,256}$	$T_{\psi_2\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$	$T_{\psi_2\psi_3\psi_4\phi_4,256}$
ν_1	48	112	32	64	128
$\tau_{K_{fa}}$	0.9558	1.8795	0.7148	1.1907	2.1063
$\hat{d}_{16}^{(t)}$	9.9741	7.7092	21.3642	17.6087	13.9983
$\hat{d}_{16}^{(e)}$	9.9970	7.6765	21.4265	17.6017	13.9554
ρ_{16}	0.9977	1.0043	0.9971	1.0004	1.0031

Tabella 3.3: Distanze soglia teoriche $\hat{d}_{16}^{(t)}$ ed empiriche $\hat{d}_{16}^{(e)}$, su 1000 simulazioni, per la rivelazione del segnale #16 utilizzando varie statistiche wavelet con soli coefficienti della prima scala a denominatore. Con ρ_{16} si indica il rapporto tra distanze teoriche ed empiriche. $\tau_{K_{fa}}$ è l'elemento separatore della regione critica dei test, mentre ν_1 sono i gradi di libertà dei numeratori; $\nu_2 = 128$ è il grado di libertà dei denominatori.

del segnale nel valore assunto dalla statistica test, come mostrato in figura 3.4.

Alla luce delle considerazioni appena svolte, utilizzando soltanto coefficienti della scala $j = 1$ per stimare la varianza del rumore nel denominatore di $T_{\phi_3,256}$, è presumibile che scompaia la discrepanza tra valore teorico e valore empirico della mediana della distanza soglia. In effetti con questa configurazione si ottiene $\rho_{16,T_{\phi_3,256}} = 1.05$, indice di un buon accordo tra la distribuzione empirica e quella teorica per la mediana della distanza soglia ottenuta con $T_{\phi_3,256}$. In tabella 3.3 vengono mostrati i risultati ottenuti, sempre sul segnale #16, da una serie di statistiche wavelet che utilizzano nei propri denominatori soltanto i coefficienti wavelet alla prima scala. Si noti come, a conferma delle considerazioni svolte in precedenza, la statistica $T_{\psi_2,256}$, con i coefficienti della seconda scala impiegati a numeratore anziché a denominatore, è caratterizzata da una distanza soglia addirittura superiore di quella che compete a $T_{\psi_3,256}$.

Come alternativa, se non si intendesse rinunciare all'uso della seconda scala per la stima della varianza del rumore, si potrebbe valutare il denominatore

della statistica test $T_{\phi_3,256}$ in un intervallo di dati precedente a quello in cui viene calcolato il numeratore. Facendo riferimento alla figura 3.4, si potrebbe valutare il denominatore nell'intervallo 90 quando il numeratore è calcolato nell'intervallo 110. In tal modo, grazie alla sua breve durata, il segnale non si manifesterebbe contemporaneamente a numeratore ed a denominatore e le assunzioni della proposizione 2.5 sarebbero soddisfatte: l'effetto di cancellazione descritto in precedenza non potrebbe aver luogo. Questo tipo di approccio potrebbe essere applicato a tutte le statistiche della sezione 2.2, ma affinché abbia senso si deve assumere che il rumore si mantenga stazionario per tempi superiori a quello che intercorre tra gli intervalli utilizzati a numeratore ed a denominatore.

Come si può dedurre dai valori riportati in tabella 3.2, i miglioramenti ottenibili con queste modifiche delle statistiche test sarebbero apprezzabili solamente utilizzando $T_{\phi_3,256}$ ed assai probabilmente risulterebbero rilevanti solo per il segnale #16. Come esempio in figura 3.5 sono riportati gli stessi grafici della figura 3.4, ma per il segnale #75 ad una distanza di 35 Kpc e per la statistica $T_{\phi_4,256}$: il segnale non fornisce contributi rilevanti alla seconda scala e la statistica test riesce a rivelarlo in questa simulazione, benché la distanza scelta sia uguale alla mediana teorica della distanza soglia.

Al di là del metodo impiegato per il loro calcolo, i valori delle mediane delle distanze soglia per i 78 segnali risultano estremamente eterogenei anche per una stessa statistica test. Perciò un commento generale su questi valori è piuttosto difficile. Nella figura 3.6 è tracciato l'istogramma dei valori trovati per alcune delle statistiche considerate, mentre in tabella 3.4 si riporta un sommario dei risultati trovati.

La distanza soglia è un parametro estremamente significativo dal punto di vista dell'applicazione che si sta trattando: permette di capire le potenzialità di rivelazione di sorgenti da parte del dispositivo sperimentale usato. Tuttavia dai valori delle distanze soglia non si ha alcuna indicazione sulla qualità del test statistico impiegato. Alcuni segnali sono intrinsecamente deboli e di difficile rivelazione anche utilizzando il metodo ottimale costituito dal test del rapporto di verosimiglianza, descritto nella sezione 3.1.1.

Applicando il test ottimale ai segnali in \mathcal{S}^* i valori mediani delle distanze soglia oscillano tra $\hat{d}_{\min} \simeq 1.4$ Kpc e $\hat{d}_{\max} \simeq 127$ Kpc; in media si ha $\bar{d} =$

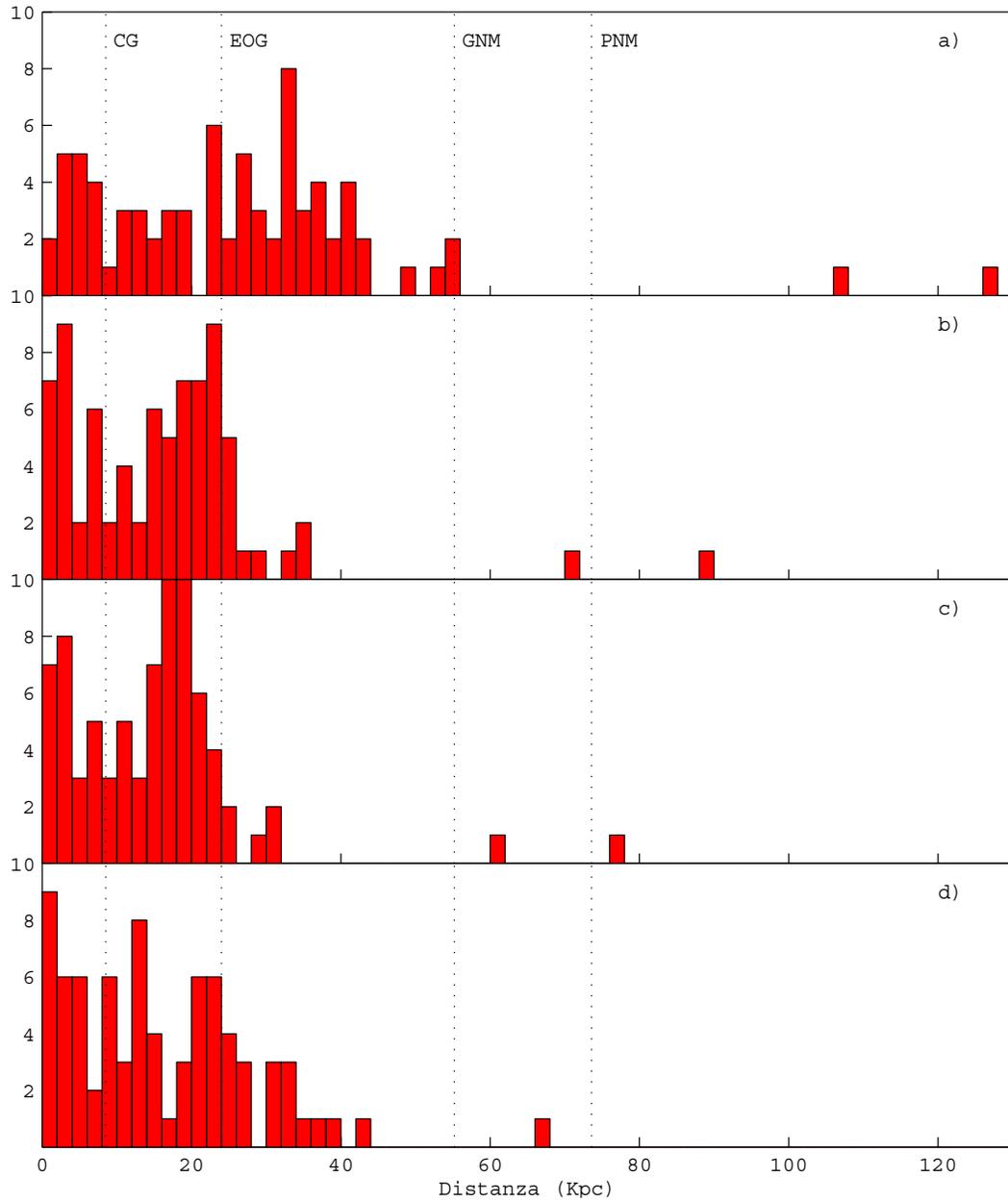


Figura 3.6: Istogrammi delle distanze soglia per i segnali di \mathcal{S}^* utilizzando: a) il test del rapporto di verosimiglianza; b) la statistica $T_{\phi_4,256}$; c) la statistica $T_{\psi_4\phi_4,256}$; d) lo *Slope Filter* in intervalli di 40 dati separati di un solo dato l'uno dall'altro. Sono riportate come linee tratteggiate alcune distanze significative: CG=centro della galassia (8.4 Kpc), EOG=estremo opposto della galassia (24 Kpc), GNM=Grande Nube di Magellano (55 Kpc), PNM=Piccola Nube di Magellano (74 Kpc).

	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$
Distribuzioni teoriche						
\hat{d}_{\min}	0.0339	0.2796	0.9013	0.2280	0.8095	0.6800
\hat{d}	1.2274	4.1870	16.3522	3.4938	14.9055	12.5545
\hat{d}_{\max}	4.6164	13.7422	89.6386	10.9236	77.4202	64.7996
Distribuzioni empiriche						
\hat{d}_{\min}	0.0339 [†]	0.2812	0.9113	0.2294	0.8189	0.6842
\hat{d}	1.1914 [†]	4.1910	16.3747	3.4906	14.9449	12.5795
\hat{d}_{\max}	4.5375	13.8809	89.3632	10.9993	77.4154	64.6379

Tabella 3.4: Valori minimi, medi, massimi e deviazione standard per le distanze soglia (in Kpc) dei 78 segnali di \mathcal{S}^* rivelati con sei statistiche diverse. I parametri utilizzati sono indicati in tabella 3.2. Il simbolo \dagger indica valori ottenuti con l'esclusione del segnale #16.

26.1 Kpc, una distanza comparabile a quella che separa la Terra dal punto più remoto della nostra galassia. In figura 3.6 è riportato l'istogramma delle mediane delle distanze soglia anche per il test del rapporto di verosimiglianza.

Per un confronto più analitico delle prestazioni di una statistica test T sul segnale i -esimo con quelle del test ottimale, si utilizza il rapporto tra le mediane delle distanze soglia corrispondenti $\varepsilon_{i,T}$, definito nell'equazione 3.11. I valori che si ottengono, per alcune delle statistiche utilizzate, sono mostrati in figura 3.7, mentre in tabella 3.5 si riportano alcune statistiche di sintesi.

Le prestazioni migliori sono ottenute dalle statistiche $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$. Solamente per 20 dei 78 segnali la statistica $T_{\psi_4\phi_4,256}$ ottiene prestazioni migliori di $T_{\phi_4,256}$, mentre solo per il segnale #39 i risultati migliori sono ottenuti da $T_{\psi_4,256}$.

Si poteva intuire a priori che $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$ avrebbero avuto prestazioni migliori delle altre statistiche wavelet sui segnali considerati. Infatti dalle conoscenze teoriche si sapeva che i segnali da rivelare, in generale, non solo in \mathcal{S}^* , sono caratterizzati prevalentemente da frequenze inferiori a 1-2 KHz; d'altra parte i coefficienti di scala per $J = 4$ agiscono proprio come filtri passa-basso con una frequenza di taglio superiore attorno a 600 Hz. Dunque risultano avvantaggiate tutte le statistiche che utilizzano questi coefficienti, e che sono caratterizzate dal pedice ϕ_4 . La statistica $T_{\psi_3\psi_4\phi_4,256}$ si comporta leggermente

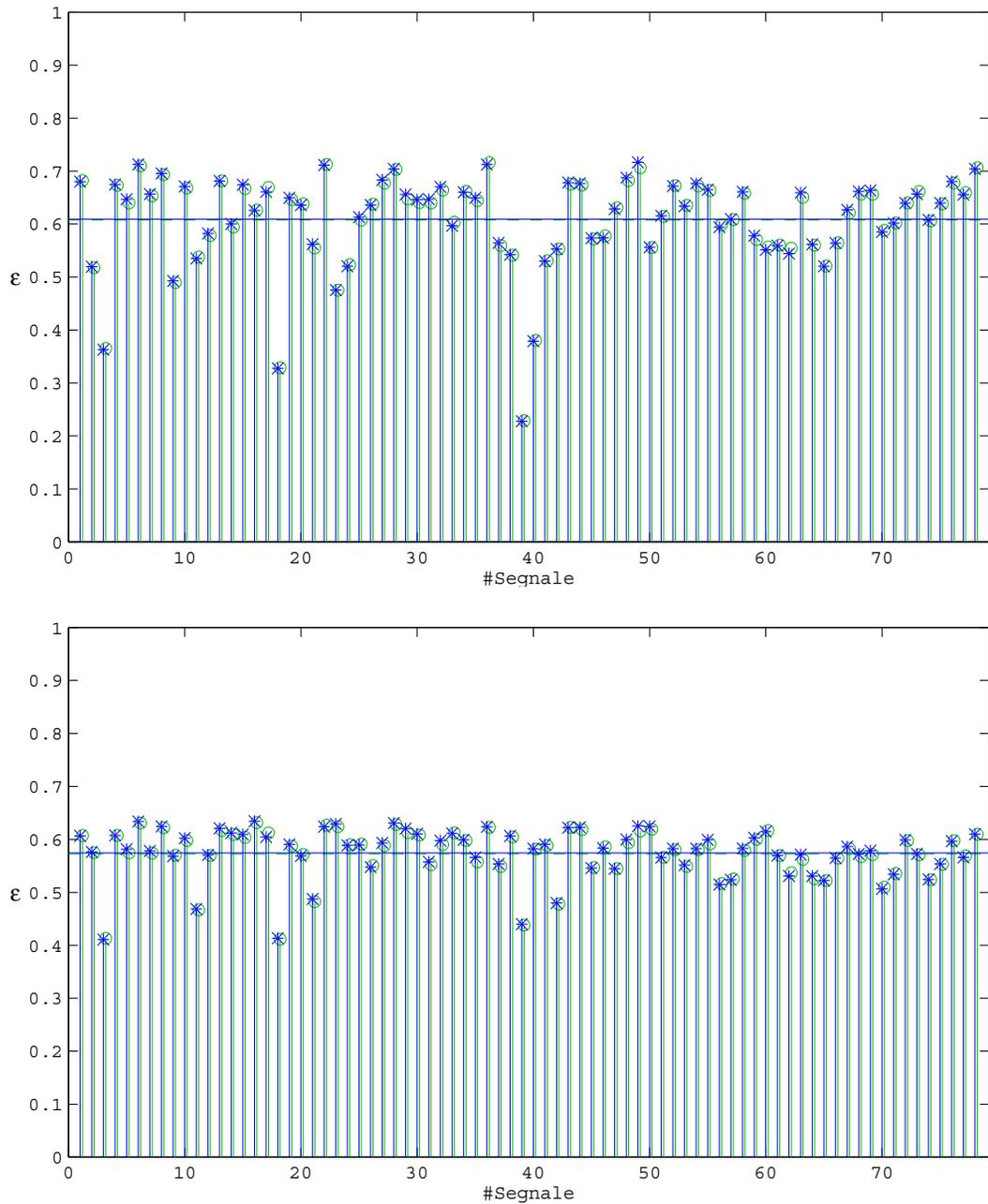


Figura 3.7: *In alto*: rapporti di prestazioni $\varepsilon_{i,T_{\phi_4,256}}$, sui 78 segnali di \mathcal{S}^* , per la statistica $T_{\phi_4,256}$. I cerchi indicano i risultati teorici, il cui valore medio è in corrispondenza dalla linea orizzontale tratteggiata; gli asterischi indicano i risultati ottenuti con 1000 simulazioni di rumore, con media segnata dalla linea orizzontale continua. I parametri impiegati sono gli stessi riportati in tabella 3.2. *In basso*: come per la figura in alto, ma per la statistica $T_{\psi_4 \phi_4,256}$.

	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$
Distribuzioni teoriche						
ε_{\min}	0.0010	0.0086	0.2287	0.0068	0.4115	0.3623
$\bar{\varepsilon}$	0.0692	0.2164	0.6078	0.1820	0.5724	0.4843
ε_{\max}	0.2013	0.5565	0.7162	0.4625	0.6315	0.5325
σ_ε	0.0522	0.1459	0.0880	0.1202	0.0475	0.0399
Distribuzioni empiriche						
ε_{\min}	0.0010 [†]	0.0087	0.2276	0.0068	0.4108	0.3623
$\bar{\varepsilon}$	0.0675 [†]	0.2163	0.6090	0.1817	0.5742	0.4855
ε_{\max}	0.1990	0.5569	0.7164	0.4650	0.6343	0.5337
σ_ε	0.0513 [†]	0.1459	0.0890	0.1202	0.0481	0.0402

Tabella 3.5: Valori minimi, medi, massimi e deviazione standard dei rapporti di prestazioni $\varepsilon_{i,T}$, nelle stesse condizioni indicate in tabella 3.2. Il simbolo † indica valori ottenuti con l'esclusione del segnale #16.

peggio delle altre due poiché utilizza anche i coefficienti wavelet alla scala $j = 3$ che agiscono sui dati come filtri passa-banda per frequenze attorno a 2.5 KHz: i risultati ottenuti mostrano che in questa banda evidentemente viene colto, di norma, più rumore che segnale.

Oltre a mostrare i più elevati valori per la media $\bar{\varepsilon}$ degli indici di prestazioni sui 78 segnali di \mathcal{S}^* , le statistiche $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$ si mostrano anche particolarmente robuste. Infatti, a parte pochissimi casi particolari, non ci sono grandi differenze tra i valori di $\varepsilon_{i,T}$ ottenuti per i vari segnali considerati e, conseguentemente, i valori di σ_ε sono limitati. Questa proprietà risulta estremamente desiderabile quando, come nel caso considerato, l'insieme \mathcal{S} dei segnali che si vorrebbe rivelare è molto esteso e poco definito. Infatti si ricorda che i segnali a cui si è potenzialmente interessati non sono affatto ristretti ai soli 78 segnali di \mathcal{S}^* .

Si fa notare che i valori di $\varepsilon_{i,T}$ trovati sono indipendenti dal valore di σ_X che è stato considerato. Infatti, tenendo conto dell'espressione 2.41 e della 3.10, si dimostra facilmente che mantenendo costante il prodotto $\sigma_X d$ il valore del parametro di non centralità δ non varia. Perciò, in questo caso, la distribuzione delle statistiche wavelet sotto ipotesi alternativa rimane immutata; dunque raddoppiando il valore di σ_X si trovano distanze soglia dimezzate. Ma la stessa identica situazione si verifica per le distanze soglia ottenute con il test

del rapporto di verosimiglianza, come si dimostra facilmente considerando la proposizione 3.1. Dal momento che in entrambi i casi le distanze soglia sono inversamente proporzionali a σ_X , i valori di $\varepsilon_{i,T}$ rimangono immutati al variare di σ_X .

La combinazione di buone prestazioni ed elevata robustezza rispetto al tipo di segnale rende le statistiche $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$ particolarmente interessanti nel contesto delle soluzioni proposte in letteratura per discriminare tra le due ipotesi del modello 2.1. In [28] viene presentato un confronto tra le statistiche test descritte nella sezione 2.1: i risultati sono riassunti in tabella 3.6.

La statistica denominata *Slope Filter*, indicata come S_n nell'equazione 2.7, emerge come la migliore: se la si impiega nelle stesse condizioni sperimentali adottate per conseguire i risultati della tabella 3.5, si ottiene addirittura $\bar{\varepsilon} = 0.79$; tuttavia questo risultato è ottenuto utilizzando una griglia di valori per la larghezza n degli intervalli di dati da analizzare: tale griglia è definita sulla base di una partizione in classi dell'insieme \mathcal{S}^* , possibile solamente grazie alla conoscenza a priori dei segnali in questo insieme. Non è affatto detto che la griglia di valori scelta sia altrettanto conveniente per segnali diversi da quelli in \mathcal{S}^* . Considerazioni del tutto analoghe possono essere svolte per i risultati quotati in tabella 3.6 per le statistiche *Offset Filter* e *Peak Correlator*, che pure mostrano elevati valori di $\bar{\varepsilon}$.

Utilizzando un singolo valore per la larghezza n dell'intervallo di dati da analizzare, lo *Slope Filter* consente di ottenere al massimo $\bar{\varepsilon} \simeq 0.62$ sui segnali di \mathcal{S}^* , come indicato in [9] e [28]. Questo valore di $\bar{\varepsilon}$ è ottenuto per $n = 40$, cioè 2 millisecondi di dati a 20 KHz, con una separazione di un solo dato tra

	BC	NF	NA	SF	OF	PC
\bar{d}	10.9	11.5	11.4	20.7	20.0	18.5
$\bar{\varepsilon}$	0.43	0.46	0.47	0.79 [‡]	0.78 [‡]	0.73 [‡]

Tabella 3.6: Medie dei rapporti di prestazioni sui segnali di \mathcal{S}^* per le statistiche test presentate in letteratura, con le stesse caratteristiche di rumore e lo stesso tasso di falsi allarmi impiegati per ottenere i risultati della tabella 3.5. Il simbolo [‡] indica risultati ottenuti per mezzo di qualche adattamento alle caratteristiche specifiche di sottoinsiemi di \mathcal{S}^* . BC=*Bin Counting*; NF=*Norm Filter*; NA=*Norm of the Autocorrelation*; SF=*Slope Filter*; OF=*Offset Filter*; PC=*Peak Correlator*.

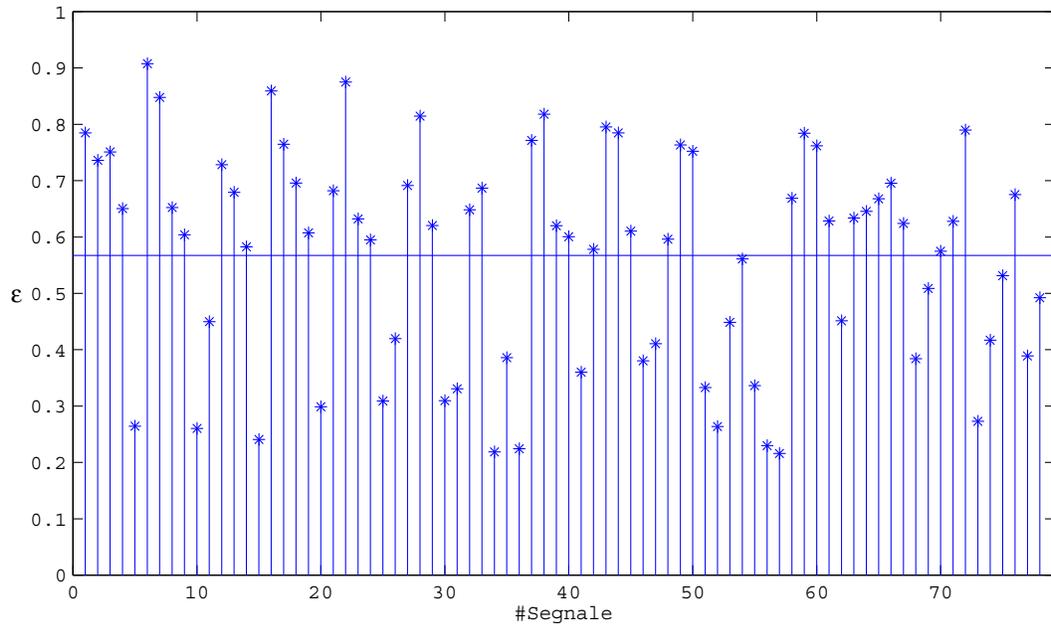


Figura 3.8: Rapporti di prestazioni $\varepsilon_{i,S_{40}}$ conseguiti sui segnali di \mathcal{S}^* dalla statistica *Slope Filter* in intervalli di 40 dati ciascuno, separati di un solo dato l'uno dall'altro. Risultati ottenuti con 500 simulazioni indipendenti di rumore senza falsi allarmi. Stesse condizioni di rumore e stesso tasso medio di falsi allarmi utilizzati per la figura 3.7.

un intervallo analizzato ed il successivo. Ancora una volta si nota un certo grado di adattamento alle caratteristiche degli specifici segnali presenti in \mathcal{S}^* , dal momento che il valore di n per lo *Slope Filter* viene scelto a posteriori, sulla base del migliore risultato ottenuto. Tuttavia in [9] viene messo in rilievo che, per questa statistica, scelte moderatamente diverse per n non comportano drastiche diminuzioni del valore di $\bar{\varepsilon}$.

Perciò si è scelto di utilizzare lo *Slope Filter* con $n = 40$ per calcolare i valori di $\varepsilon_{i,S_{40}}$ mostrati in figura 3.8. A tale scopo sono state impiegate 500 simulazioni della serie di rumore, indipendenti e prive di falsi allarmi. Il valore medio che si è ottenuto risulta pari a $\bar{\varepsilon} \simeq 0.57$ nel caso in cui si considerino le mediane delle distribuzioni empiriche delle distanze soglia. In [9] invece sono state usate le medie delle distribuzioni empiriche per ottenere il valore $\bar{\varepsilon} \simeq 0.62$; in effetti, utilizzando la media al posto della mediana, anche per le simulazioni utilizzate nel presente lavoro si ottiene $\bar{\varepsilon} \simeq 0.60$, in buon accordo con il valore

riportato in letteratura. D'altra parte, come discusso in precedenza, si ritiene la mediana un indice più robusto rispetto alla media per la locazione di una distribuzione generica.

Le statistiche $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$, per le quali si ha rispettivamente $\bar{\varepsilon} \simeq 0.61$ e $\bar{\varepsilon} \simeq 0.57$, risultano pertanto avere delle prestazioni medie ben comparabili con quelle dello *Slope Filter*, garantendo nel contempo una robustezza assai maggiore rispetto al tipo di segnale considerato. Infatti, dal confronto delle figure 3.8 e 3.7, appare evidente che lo *Slope Filter* risulta molto più disomogeneo nei risultati che riesce ad ottenere su segnali diversi: per questa statistica test risulta $\sigma_\varepsilon \simeq 0.19$ a fronte di $\sigma_\varepsilon \simeq 0.09$ e $\sigma_\varepsilon \simeq 0.05$ ottenuti da $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$ rispettivamente. Inoltre per lo *Slope Filter* si ha un valore $\varepsilon_{\min} \simeq 0.21$ inferiore rispetto alle altre due statistiche.

3.2.1 Dipendenza dall'ampiezza α del test

Si è provato a valutare l'influenza della scelta dell'ampiezza α sulle prestazioni dei test di ipotesi proposti nella sezione 2.2. Perciò si è fissato $\alpha = 10^{-6}$, un valore 16 volte inferiore a quello usato per ottenere i risultati riportati in tabella 3.5; gli altri parametri sono rimasti invariati. Con questo valore di α in media solo un test ogni milione da luogo ad un falso allarme; con una frequenza di campionamento di 20 KHz ed una separazione di 16 dati tra intervalli consecutivi analizzati, dall'equazione 2.46 si ricava un tasso atteso di falsi allarmi pari a $\kappa_{fa} = (1/800) \text{ sec}^{-1}$, ma si deve ricordare che i falsi allarmi non risultano distribuiti omogeneamente nel tempo, come si è discusso nella sezione 2.2.2. I risultati ottenuti con questo cambiamento di α sono riassunti in tabella 3.7 e sono stati ricavati utilizzando la distribuzione teorica sotto ipotesi alternativa delle statistiche considerate.

È evidente una diminuzione dei valori delle distanze soglia per tutte le statistiche, indice di un generale peggioramento delle prestazioni. Tuttavia, diminuendo il tasso atteso di falsi allarmi κ_{fa} , diminuiscono anche le distanze soglia ottenute con il test del rapporto di verosimiglianza: come conseguenza si hanno indici di prestazioni ε praticamente identici a quelli corrispondenti ad un tasso di falsi allarmi 16 volte superiore: si confrontino i valori della tabella 3.7 con quelli della tabella 3.5.

Dunque le statistiche proposte si mantengono assai valide anche nel caso in

	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$
$\tau_{\kappa_{\text{fa}}}$	0.5116	0.3410	0.3410	0.6711	0.5116	0.8252
\hat{d}_{\min}	0.0307	0.2523	0.8132	0.2071	0.7340	0.6184
\hat{d}	1.1128	3.7782	14.7554	3.1741	13.5137	11.4189
\hat{d}_{\max}	4.1851	12.4009	80.8835	9.9246	70.1966	58.9348
ε_{\min}	0.0010	0.0086	0.2294	0.0068	0.4148	0.3664
$\bar{\varepsilon}$	0.0698	0.2171	0.6099	0.1838	0.5771	0.4899
ε_{\max}	0.2030	0.5585	0.7188	0.4672	0.6367	0.5386
σ_ε	0.0527	0.1464	0.0883	0.1215	0.0479	0.0404

Tabella 3.7: Risultati teorici analoghi a quelli riportati nelle tabelle 3.4 e 3.5, ma variando le soglie $\tau_{\kappa_{\text{fa}}}$ in modo da ridurre il tasso medio di falsi allarmi a $\kappa_{\text{fa}} = (1/800) \text{ sec}^{-1}$: l'ampiezza dei singoli test si riduce da $\alpha = 1.6 \cdot 10^{-5}$ a $\alpha = 10^{-6}$ in questo caso. Gli altri parametri sono lasciati invariati.

cui nella pratica sperimentale si fosse costretti a ridurre il tasso di falsi allarmi per riuscire a gestire la trattazione successiva di questo tipo di eventi.

3.2.2 Dipendenza dalla larghezza n dell'intervallo

Uno dei parametri che definiscono le statistiche della sezione 2.2 è la larghezza n dell'intervallo di dati su cui si effettua un singolo test di ipotesi. Si è pensato di studiare quali siano le conseguenze della scelta operata per n sulle prestazioni ottenute dai test di ipotesi che utilizzano queste statistiche.

A tale scopo si è scelto di considerare, a titolo di esempio, una coppia di segnali di \mathcal{S}^* : il segnale #39 è stato selezionato poiché per esso la statistica $T_{\phi_4,256}$ consegue prestazioni nettamente peggiori rispetto a tutti gli altri segnali; il segnale #21 è stato considerato interessante per questo studio poiché è uno di quelli con una durata nettamente superiore ai 10 millisecondi ipotizzati a priori per fissare la scelta $n = 256$.

Per una serie di scelte del parametro n sono stati quindi calcolati i valori di ε_{39,T_n} e di ε_{21,T_n} , grazie alla conoscenza delle distribuzioni teoriche delle statistiche wavelet T_n sotto ipotesi alternativa ottenute dalla proposizione 2.5.

Si è continuato a considerare lo stesso tasso di falsi allarmi $\kappa_{\text{fa}} = (1/50) \text{ sec}^{-1}$ e la stessa varianza del rumore $\sigma_X^2 = 1.6 \cdot 10^{-41}$ utilizzati per ricavare i risultati della tabella 3.5. Tuttavia i valori di soglia $\tau_{\kappa_{\text{fa}}}$ che delimitano le

regioni critiche delle varie statistiche non sono rimasti gli stessi della tabella 3.1. Infatti variando n cambiano anche i numeri ν_1 e ν_2 di coefficienti impiegati a numeratore ed a denominatore nelle varie statistiche test: per conseguenza variano anche i gradi di libertà delle distribuzioni di Fisher che competono a tali statistiche sotto ipotesi nulla e quindi le soglie $\tau_{\kappa_{fa}}$ devono essere ricalcolate per ogni valore di n .

Nel caso del segnale #39 le migliori prestazioni, misurate sulla base di ε_{39, T_n} , si ottengono per $n = 64$ sia utilizzando $T_{\phi_4, n}$ che $T_{\psi_4 \phi_4, n}$; anche per $T_{\psi_4, n}$, la statistica che solo per questo segnale risulta preferibile alle altre due, il valore ottimale risulta essere $n = 64$. Se vengono fissati valori superiori per n , come ad esempio $n = 256$ utilizzato in precedenza, si ha un lento decadimento delle prestazioni, come mostrato nella figura 3.9. Se invece si scelgono valori di n inferiori a 64 il peggioramento risulta assai più brusco.

Lo stesso tipo di comportamento si osserva per il segnale #21: anche in questo caso si trova un massimo di ε_{21, T_n} in corrispondenza di $n = 64$ ed una pendenza maggiore a sinistra di tale massimo, per tutte e tre le statistiche T_n considerate in precedenza. Tuttavia per questo segnale le prestazioni migliori si ottengono in corrispondenza di $n \geq 400$, come mostrato sempre nella figura 3.9. Osservando l'andamento temporale del segnale #21, riportato in figura 3.10, si capiscono le ragioni di questo comportamento: con una larghezza $n \geq 400$ i due picchi principali del segnale cadono all'interno di uno stesso intervallo di dati su cui viene effettuato il test di ipotesi.

In entrambi i casi considerati l'impiego del valore $n = 256$ comporta, per tutte e tre le statistiche $T_{\phi_4, n}$, $T_{\psi_4 \phi_4, n}$, $T_{\psi_4, n}$, una perdita del 10-15% su ε_{i, T_n} rispetto ad una scelta ottimale di n .

La presenza, per entrambi i segnali, di un massimo per ε_{i, T_n} in corrispondenza di $n = 64$ potrebbe suggerire di scegliere valori relativamente bassi per n , adattando in tal modo le statistiche test alla ricerca dei singoli picchi nel segnale. Tuttavia il caso del segnale #21 evidenzia chiaramente i vantaggi dell'utilizzo di intervalli sufficientemente larghi da includere, con un unico posizionamento, molteplici picchi nel segnale. Si deve poi considerare che questi vantaggi dovrebbero essere ancor più rilevanti nel caso di picchi più ravvicinati di quelli del segnale #21. In ogni caso non è lecito svolgere a posteriori questo genere di considerazioni sulla scelta del valore da adottare per n : infatti tale

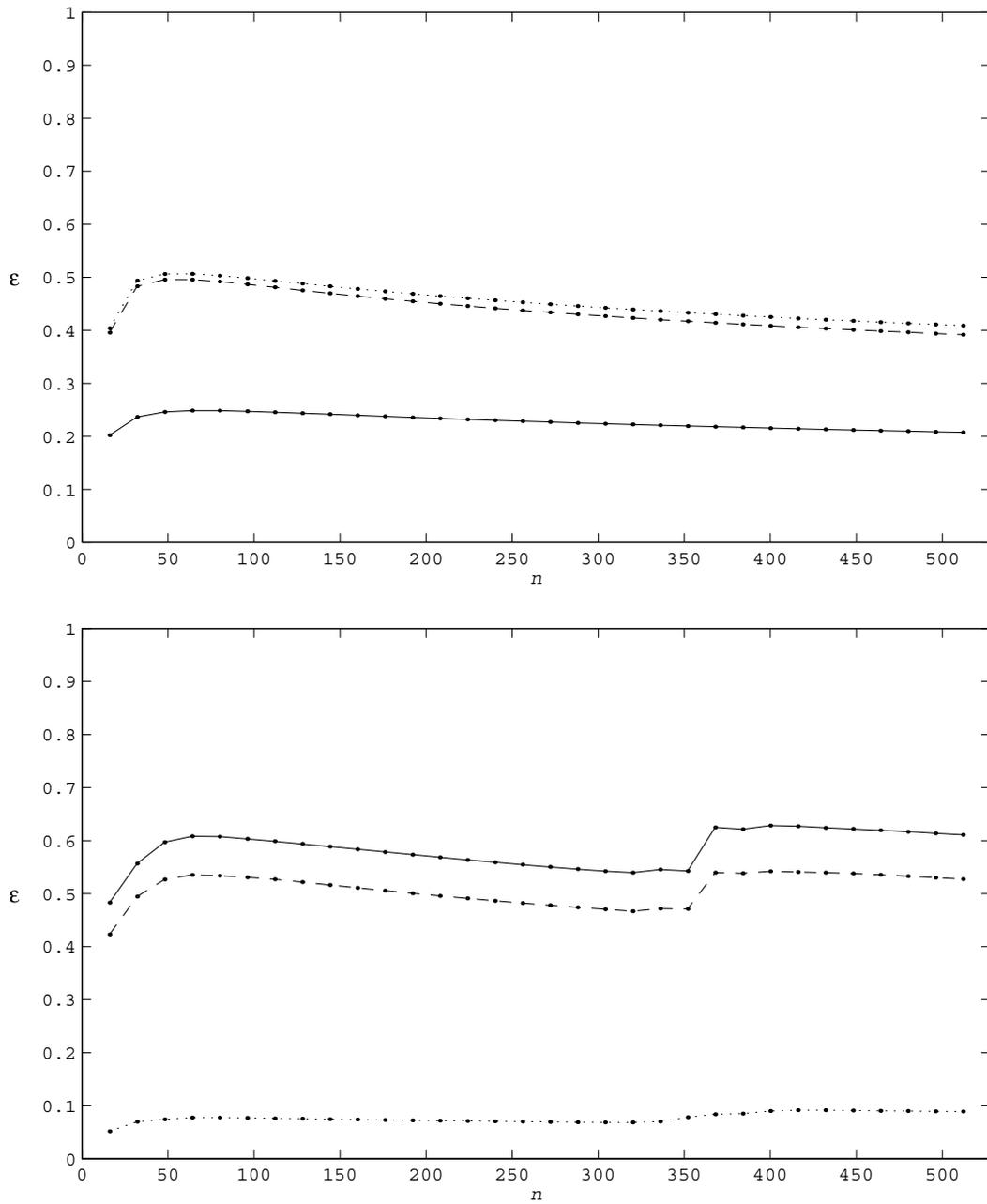


Figura 3.9: *In alto*: rapporto di prestazioni ε_{39, T_n} sul segnale #39 in funzione della larghezza n dell'intervallo di dati considerato in ogni singolo test, con intervalli separati di 16 dati l'uno dall'altro. I risultati si basano sulle distribuzioni teoriche e si riferiscono alle statistiche $T_{\phi_4, n}$ (linea continua), $T_{\psi_4, n}$ (linea punteggiata) e $T_{\psi_4\phi_4, n}$ (linea tratteggiata). *In basso*: grafici come nel riquadro superiore, ma per il segnale #21.

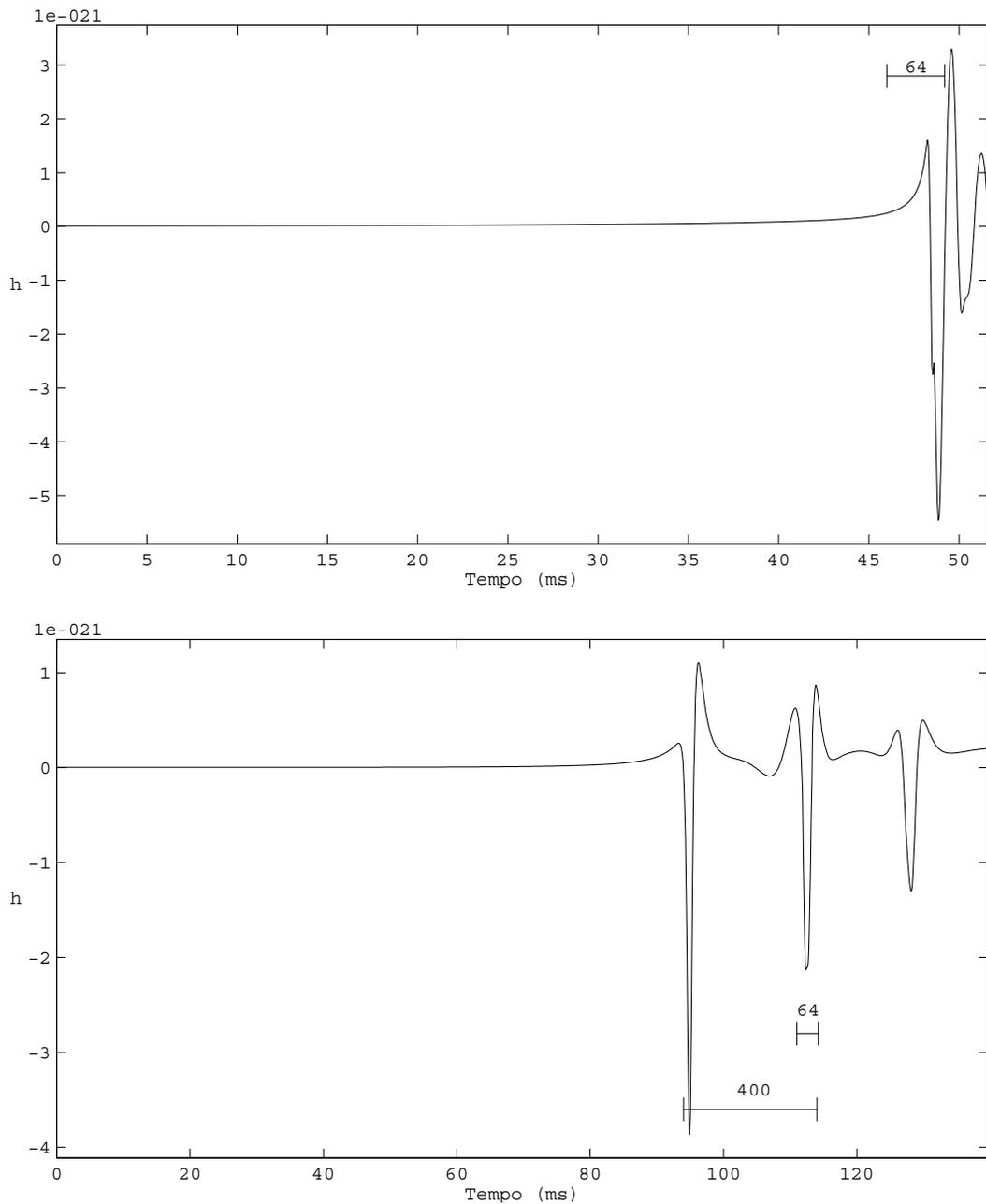


Figura 3.10: Andamento temporale del segnale #39 ad una distanza di 13 Kpc (*in alto*) e del segnale #21 a 48 Kpc (*in basso*); le distanze scelte sono le distanze soglia per il test del rapporto di verosimiglianza su questi specifici segnali. In entrambi i grafici è mostrata l'estensione di un intervallo di 64 dati, pari a 3.2 millisecondi per una frequenza di campionamento di 20 KHz. Nel grafico inferiore si mostra anche un intervallo di 400 dati (20 ms).

scelta deve essere operata in realtà solo sulla base delle conoscenze a priori sul tipo di segnali da rivelare, come si è fatto quando si è fissato $n = 256$.

Infine si rendono necessarie alcune osservazioni sull'andamento temporale del segnale #39, mostrato in figura 3.10. Come regola generale si è scelto di non fare alcun tipo di estrapolazione sui segnali forniti in [52], utilizzando la loro intera durata come intervallo temporale esteso su cui effettuare la trasformata wavelet discreta, seguendo le indicazioni della sezione 2.2.3. Tuttavia, per evitare ogni possibile effetto di bordo nei coefficienti così ottenuti, non sono stati considerati i risultati delle statistiche test nei primi e negli ultimi tre intervalli di analisi di larghezza n necessari per coprire l'intero intervallo esteso. Con intervalli separati di 16 dati l'uno dall'altro ed una frequenza di campionamento di 20 KHz, ciò corrisponde a trascurare circa 2 millisecondi all'inizio e alla fine dei segnali forniti da [52]. Perciò per il segnale #39 non vengono considerati gli effetti delle ultime due oscillazioni mostrate in figura 3.10: questo effetto spiega in parte le non eccelse prestazioni ottenute su questo segnale dalla statistica test $T_{\phi_4, n}$. Ma soprattutto si spiega la corrispondenza tra il valore ottimale $n = 64$ e le dimensioni di un singolo picco, nonostante la presenza di una serie di picchi ravvicinati nel segnale che dovrebbero garantire prestazioni migliori per $n > 64$.

Naturalmente il problema appena descritto non si presenterà nell'applicazione reale, dal momento che esso è frutto esclusivamente delle difficoltà dei modelli teorici usati in [52] nel calcolare la parte finale di alcuni dei segnali in \mathcal{S}^* . Pertanto ci si aspetta, in generale, che alcuni dei segnali in \mathcal{S}^* potrebbero fornire valori più elevati di $\varepsilon_{i, T}$ se fossero definiti fino alla fine del fenomeno transiente.

3.2.3 Potenza dei test wavelet

Nella parte iniziale di questo capitolo sono state discusse le motivazioni che hanno portato a valutare le prestazioni dei test di ipotesi proposti in questa tesi sulla base di parametri diversi dalla potenza π , molto più usuale in un contesto statistico. Essenzialmente si poneva un problema di confrontabilità con i risultati già presenti in letteratura per questo specifico contesto applicativo.

Sono invece condivise da tutti i possibili criteri di valutazione delle prestazioni dei test le difficoltà che derivano dalla mancanza di una definizione ben

posta per l'insieme \mathcal{S} dei segnali eventualmente presenti sotto l'ipotesi alternativa H_1 . Utilizzando come criterio quello della distanza soglia descritto in precedenza si è cercato di aggirare queste difficoltà restringendo l'attenzione ad un ben definito sottoinsieme di segnali $\mathcal{S}^* \subset \mathcal{S}$. Si può pensare di ricorrere ad una soluzione analoga anche qualora si decidesse di adottare come criterio di valutazione la potenza dei test di ipotesi; in questo caso, per ogni test sotto esame, si ricava la potenza $\Pr(H_1|H_1, s)$ conseguita su ciascun segnale s di uno specifico sottoinsieme $\mathcal{S}' \subset \mathcal{S}$. Eventualmente la potenza complessiva su \mathcal{S}' può essere calcolata per mezzo dell'espressione 3.1 attribuendo una probabilità $\Pr(s)$ a ciascun segnale $s \in \mathcal{S}'$.

D'altra parte l'insieme \mathcal{S}^* utilizzato in precedenza non pare essere una buona scelta per l'insieme \mathcal{S}' su cui valutare la potenza dei test. Infatti i segnali in \mathcal{S}^* sono tutti normalizzati per una distanza di 1 Kpc mentre si è visto che, anche per il test del rapporto di verosimiglianza, le distanze soglia che si ottengono per questi segnali sono estremamente eterogenee: ciò vuol dire che in \mathcal{S}^* i segnali sono intrinsecamente molto diversi per intensità l'uno dall'altro. Sarebbe invece auspicabile utilizzare un insieme \mathcal{S}' più omogeneo per confrontare le potenze ottenute sui segnali che vi appartengono.

Si è quindi scelto di definire l'insieme \mathcal{S}' come l'insieme dei segnali in \mathcal{S}^* normalizzati ciascuno per la mediana della corrispondente distanza soglia ottenuta col test del rapporto di verosimiglianza; ciò significa che tutti i segnali in \mathcal{S}' sono rivelati con la stessa identica probabilità dal test del rapporto di verosimiglianza: per definizione tale probabilità è del 50%. Conseguentemente per un qualsiasi test di rivelazione ed un qualsiasi segnale in \mathcal{S}' si otterrà una potenza sicuramente inferiore al valore limite 0.5.

In tabella 3.8 si presentano alcuni dati significativi per le potenze ottenute da alcune statistiche wavelet sui segnali di \mathcal{S}' ; in figura 3.11 invece vengono mostrate le potenze ottenute per ciascuno dei segnali in \mathcal{S}' dalle statistiche $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$. Si noti come il valore medio π riportato in tabella 3.8 corrisponda alla potenza complessiva dei test sotto esame sulla globalità dell'insieme \mathcal{S}' , ammettendo che i segnali che vi appartengono siano a priori equiprobabili.

Le considerazioni che si possono trarre da questo criterio di confronto delle prestazioni dei test wavelet proposti non sono molto diverse da quelle svolte sulla base del criterio delle distanze soglia e dei rapporti di prestazioni. Infatti

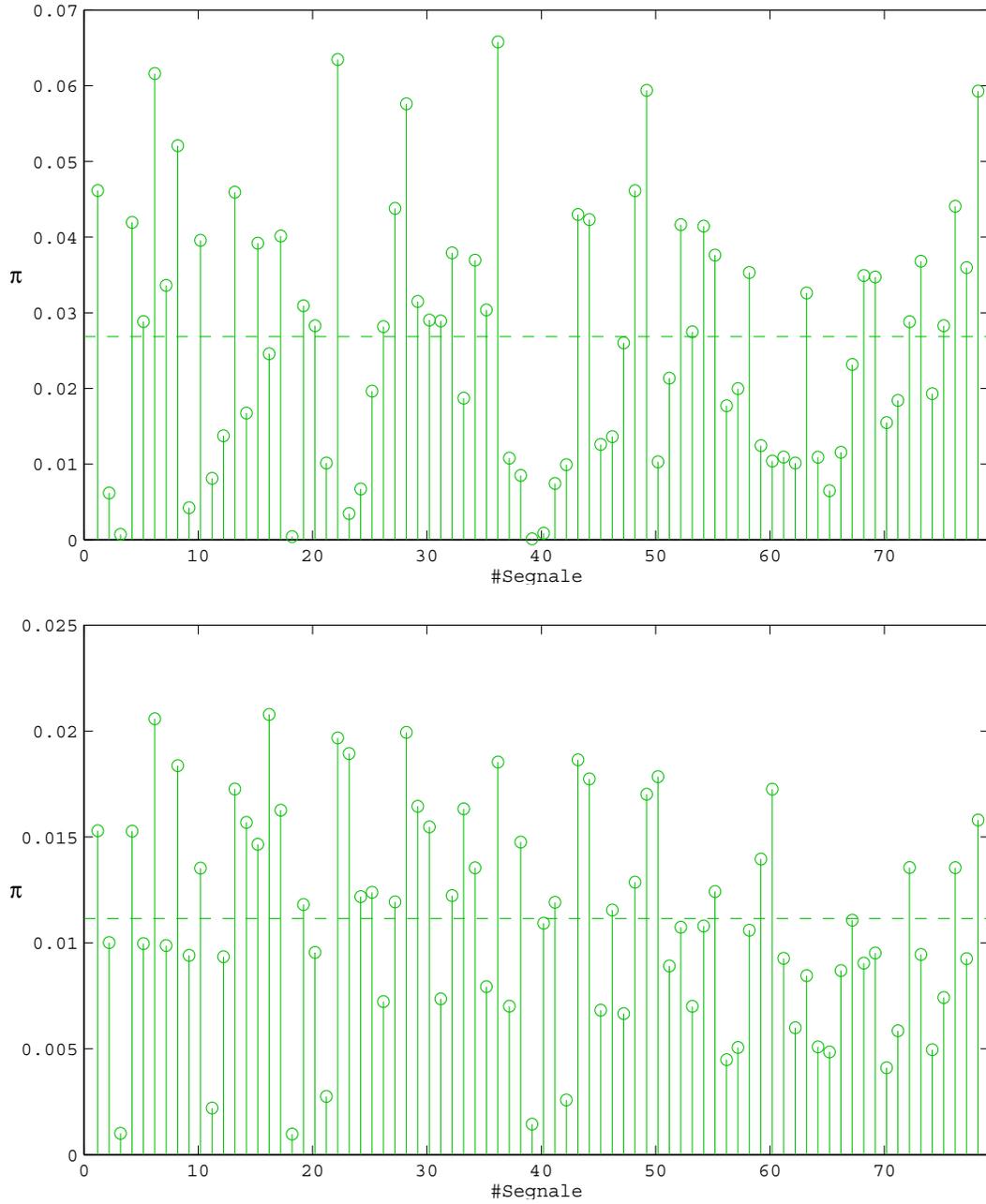


Figura 3.11: Valori delle potenze conseguite sui segnali di \mathcal{S}' dai test che utilizzano la statistica $T_{\phi_4,256}$ (in alto) e la statistica $T_{\psi_4\phi_4,256}$ (in basso). Le linee orizzontali indicano i rispettivi valori medi. Per tutti i segnali di \mathcal{S}' la potenza ottenuta col test del rapporto di verosimiglianza è pari a 0.5.

	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$
π_{\min}	$1.60 \cdot 10^{-5}$	$1.61 \cdot 10^{-5}$	$9.90 \cdot 10^{-5}$	$1.60 \cdot 10^{-5}$	$9.81 \cdot 10^{-4}$	$3.61 \cdot 10^{-4}$
π	$2.13 \cdot 10^{-5}$	$6.17 \cdot 10^{-4}$	$2.68 \cdot 10^{-2}$	$1.27 \cdot 10^{-4}$	$1.12 \cdot 10^{-2}$	$2.42 \cdot 10^{-3}$
π_{\max}	$5.60 \cdot 10^{-5}$	$1.04 \cdot 10^{-2}$	$6.58 \cdot 10^{-2}$	$1.73 \cdot 10^{-3}$	$2.08 \cdot 10^{-2}$	$4.27 \cdot 10^{-3}$
σ_π	$7.42 \cdot 10^{-6}$	$1.63 \cdot 10^{-3}$	$1.69 \cdot 10^{-2}$	$2.46 \cdot 10^{-4}$	$5.09 \cdot 10^{-3}$	$1.06 \cdot 10^{-3}$

Tabella 3.8: Valori minimi, medi, massimi e deviazioni standard per le potenze conseguite da una serie di test wavelet su 78 segnali per i quali il test del rapporto di verosimiglianza ha una potenza pari a 0.5.

anche in termini di potenza le statistiche $T_{\phi_4,256}$ e $T_{\psi_4\phi_4,256}$ risultano di gran lunga le migliori; il segnale #39 risulta l'unico per cui la potenza più elevata sia ottenuta con la statistica $T_{\psi_4,256}$, mentre solo per una quindicina di segnali $T_{\psi_4\phi_4,256}$ è più potente di $T_{\phi_4,256}$. D'altra parte $T_{\psi_4\phi_4,256}$ si propone come la statistica più robusta esibendo il valore più elevato di π_{\min} . Inoltre dal confronto della figura 3.11 con la figura 3.7 si evidenziano relazioni d'ordine tra le potenze π_i dei segnali in \mathcal{S}' analoghe alle relazioni d'ordine tra i rapporti di prestazioni ε_i dei segnali in \mathcal{S}^* .

In definitiva dunque l'utilizzo del classico criterio del confronto della potenza tra test di ipotesi diversi non sembra portare ad alcuna rilevante modifica delle valutazioni sui test wavelet svolte sulla base del criterio dei rapporti di prestazioni, criterio introdotto *ad hoc* nella specifica applicazione trattata.

Un altro indice delle prestazioni, molto utilizzato nel contesto della rivelazione di segnali, è il rapporto segnale rumore. Esistono varie possibili definizioni di questa quantità; in questo lavoro si è scelto di definire il rapporto segnale rumore RSR come il rapporto tra media e standard deviation della distribuzione, sotto ipotesi alternativa, della statistica W utilizzata nel test del rapporto di verosimiglianza: si faccia riferimento dunque alla sezione 3.1.1. Con questa definizione, sfruttando la proposizione 3.1, si ha:

$$RSR = \|\tilde{s}\| = \sqrt{\sum_k |\tilde{s}_k|^2}, \quad (3.14)$$

dove \tilde{s} è ottenuto a partire dal segnale originario s per mezzo della relazione $\tilde{s} = \Gamma s$; la matrice Γ è tale che $\Gamma'\Gamma = \Sigma^{-1}$, indicando con Σ la matrice di varianza e covarianza del rumore.

Dal momento che la trasformazione di s in \tilde{s} è lineare e che i segnali gravitazionali hanno ampiezza inversamente proporzionale alla distanza d da cui provengono, si conclude che anche il rapporto segnale rumore RSR , per un dato tipo di segnale, è inversamente proporzionale a d . D'altra parte i segnali di \mathcal{S}' , che hanno forme diverse tra loro, provengono tutti da distanze pari alle rispettive distanze soglia per il test del rapporto di verosimiglianza: quindi, per definizione, tutti i segnali in \mathcal{S}' hanno lo stesso rapporto segnale rumore RSR_o . Grazie al corollario 3.2 si calcola che, con il tasso di falsi allarmi scelto $\kappa_{fa} = (1/50) \text{ sec}^{-1}$, vale $RSR_o \simeq 4.75$.

Tenendo conto delle considerazioni appena svolte, gli indici di prestazioni $\varepsilon_{i,T}$ utilizzati nelle sezioni precedenti possono essere interpretati anche come i reciproci della frazione di RSR_o necessaria alla statistica T per rivelare l' i -esimo segnale con il 50% di probabilità. Ad esempio se $\varepsilon_{i,T} = 0.6$ allora alla statistica T occorre un rapporto segnale rumore pari a $RSR = (1/0.6) RSR_o \simeq 1.7 RSR_o \simeq 7.9$ per avere il 50% di probabilità di rivelare il segnale $\#i$.

3.3 Prestazioni nel caso di rumore correlato

Nella sezione 2.3 sono state proposte alcune metodologie per consentire l'utilizzo dei test di ipotesi della sezione 2.2 anche in presenza di un rumore correlato. Si è quindi ritenuto opportuno valutare le prestazioni di tali test di ipotesi con un modello per la struttura di correlazione del rumore più realistico rispetto all'assunzione di rumore bianco.

Per tale scopo si è utilizzato la trasformazione di *whitening* definita dall'equazione 2.50. Questo approccio è stato preferito agli altri descritti nella sezione 2.3 per le sue caratteristiche di maggiore semplicità ed universalità di utilizzo. Tuttavia questa scelta non indica affatto una minore validità degli altri metodi proposti, per i quali anzi sarà necessario un approfondimento degli studi in futuro.

La trasformazione 2.50 richiede la conoscenza della densità di potenza spettrale $S_f^{(X)}$ della serie di rumore. Inizialmente è stata considerata la situazione ideale in cui $S_f^{(X)}$ è effettivamente nota a priori. È stata utilizzata la densità di potenza spettrale mostrata nel riquadro superiore della figura 3.12, che rappresenta un modello teorico dello spettro di rumore che dovrebbe essere

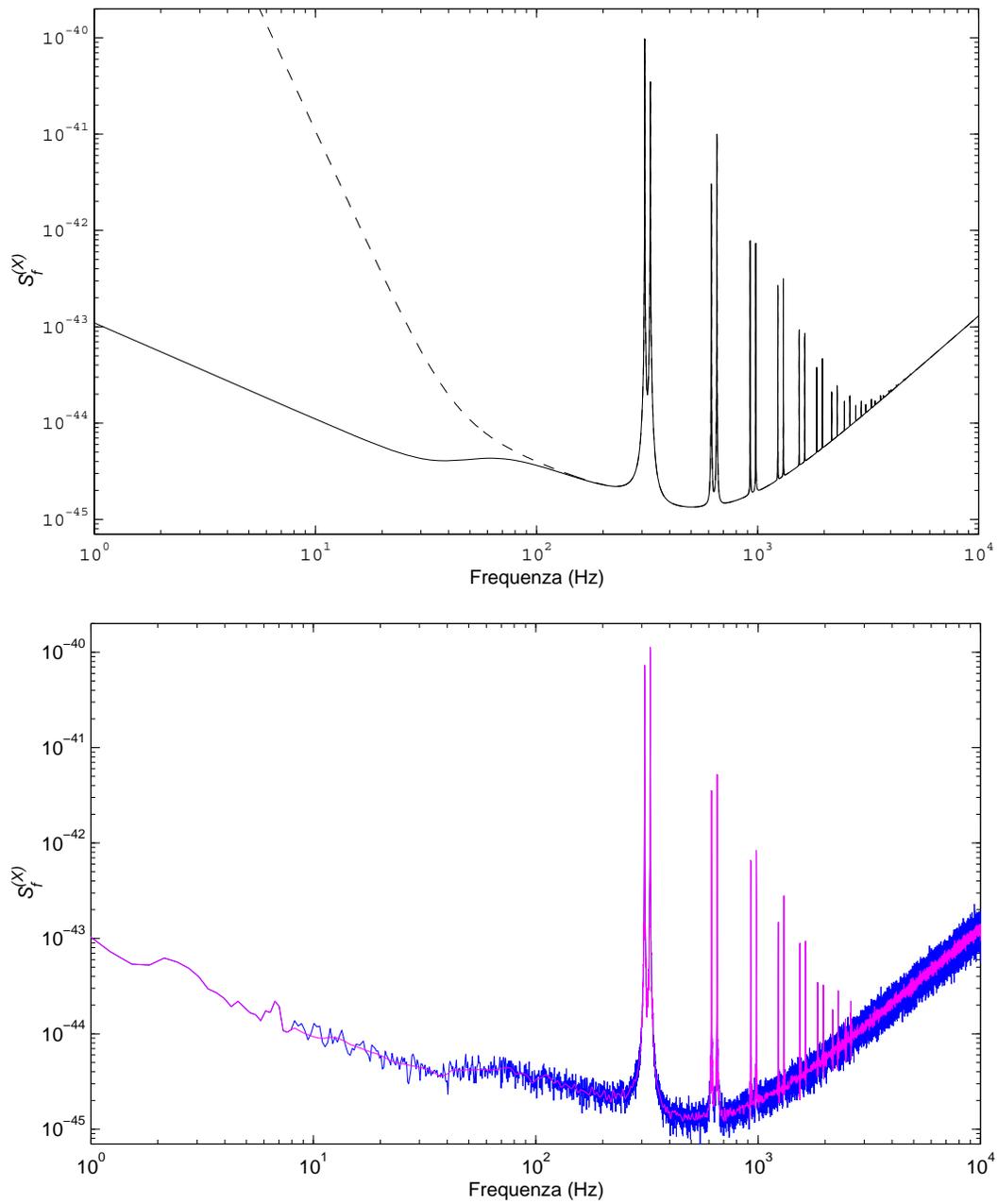


Figura 3.12: *In alto*: densità di potenza spettrale teorica per il rumore dell'interferometro VIRGO; la linea tratteggiata si riferisce al rumore prima del filtraggio di “pre-whitening”. *In basso*: stima col metodo di Welch, su una sequenza di 2^{21} dati di rumore generati con la densità mostrata in alto, della densità di potenza spettrale (linea blu). In magenta è mostrata anche la stima che si ottiene dalla precedente filtrandola con la tecnica del *wavelet shrinkage*.

prodotto dall'apparato sperimentale impiegato nel progetto VIRGO. I parametri utilizzati sono quelli forniti in [13]. D'altra parte è noto che i dati di VIRGO dovranno essere sottoposti ad un filtraggio passa-alto analogico prima di essere digitalizzati: questa operazione di “*pre-whitening*” è necessaria al fine di ridurre il *range dinamico* della serie storica che altrimenti sarebbe dominato dal contributo di rumore di bassa frequenza: infatti un *range* dinamico esteso introduce un errore di digitalizzazione più elevato, a parità di numero di *bit* utilizzati per rappresentare i dati. Per simulare il processo di *pre-whitening* appena descritto si è fatto ricorso ad un filtro passa-alto di Butterworth di ordine 2 e frequenza di taglio di 56 Hz, [37].

La conoscenza dell'andamento di $S_f^{(X)}$ è stata assunta a meno di una costante moltiplicativa incognita, eventualmente variabile nel tempo: è compito dei denominatori delle statistiche della sezione 2.2 stimare il valore di tale costante. In pratica si è fissato questo valore in modo che sia $\min_f \{S_f^{(X)}\} = 1.3441 \cdot 10^{-45} \text{ Hz}^{-1}$, non molto diverso dal valore costante che corrisponde al caso di rumore bianco considerato in precedenza, $S_f^{(X)} = 1.6 \cdot 10^{-45} \text{ Hz}^{-1}$.

I 78 segnali di \mathcal{S}^* sono stati sottoposti prima al pre-filtraggio alle basse frequenze e quindi alla trasformazione 2.50. Per i segnali così modificati sono state ricalcolate le distanze soglia sia per il test di massima verosimiglianza che per i test wavelet già impiegati nella sezione 3.2, utilizzando le stesse distribuzioni teoriche indicate dalla proposizione 2.5. Infatti il modello 2.51, nel quale figurano i segnali trasformati, gode delle stesse proprietà del modello 2.1 con rumore bianco, pur essendo equivalente al modello 2.1 con rumore correlato.

Dato che l'assunzione sulla conoscenza della densità di potenza spettrale del rumore $S_f^{(X)}$ è assai poco realistica, si è cercato una conferma dei risultati trovati per via teorica con tale assunzione ricorrendo a simulazioni. La procedura seguita è di una certa complessità: per ciascuna delle 1000 simulazioni utilizzate sono state eseguite le seguenti operazioni:

1. È stata generata una serie di rumore correlato di 2^{21} dati consecutivi nel dominio dei tempi, utilizzando la densità di potenza spettrale teorica $S_f^{(X)}$ e sfruttando la rappresentazione di Cramer del rumore, [7], adattata

al caso discreto:

$$X_k = \frac{1}{2\pi} \sum_f \sqrt{S_f^{(X)}} W_f \exp [i2\pi f k], \quad (3.15)$$

dove il peso W_f è ottenuto dalla trasformata di Fourier discreta di una realizzazione di una serie di rumore bianco a varianza unitaria. Dalla serie di dati X_k ottenuta è stata quindi estratta una serie di 2^{16} dati consecutivi.

2. Le due serie di rumore, di 2^{21} e 2^{16} dati, sono state sottoposte al filtraggio di *pre-whitening* descritto in precedenza. Sui segnali invece questa operazione è stata effettuata una sola volta per tutte le simulazioni. Infatti, trattandosi di un'operazione lineare, è possibile applicare l'operazione di *pre-whitening* separatamente al rumore ed al segnale.
3. La serie di 2^{21} dati di rumore è stata utilizzata per stimare lo spettro del rumore strumentale ottenuto dopo il filtraggio preliminare alle basse frequenze. Molte sono le metodologie che possono essere impiegate per questa stima, si veda ad esempio [40] o [38]. È assai probabile che molte di queste metodologie risulterebbero adeguate per i nostri scopi, ma si è voluto cogliere l'occasione per proporre una nuova procedura di stima che sfrutta, anche in questo contesto, lo strumento delle wavelets.

In primo luogo è stata ottenuta una stima spettrale col classico metodo di Welch, con finestre di 2^{16} dati non sovrapposte; per dettagli su questo metodo si rimanda a [40] o [38]. D'altra parte è noto a priori che lo spettro da stimare è in genere piuttosto "liscio" (regolare), ad eccezione di ristretti intervalli attorno ad alcune ben note frequenze di risonanza caratteristiche delle sospensioni dello strumento. Perciò si è pensato di sottoporre la stima di Welch ad un algoritmo di "lisciamento" come il *wavelet shrinkage* proposto da Donoho e Johnstone, [17], [18].

Si è ritenuto più opportuno applicare il *wavelet shrinkage* su una stima del log-spettro, in modo da evitare che nel processo si originassero valori negativi inaccettabili per la stima spettrale. Perciò è stata calcolata la trasformata wavelet discreta del logaritmo della stima di Welch, osservando che tale trasformata è applicabile su serie di dati indicizzati da

una frequenza esattamente come nel caso, già discusso, di serie indicizzate da un tempo. Sono quindi stati posti uguali a zero tutti i coefficienti della trasformata che non superavano in valore assoluto la “soglia universale” $\lambda = \sigma_c \sqrt{2 \log_2(n)}$, definita in [19], dove σ_c è una stima della deviazione standard dei coefficienti e n il numero di dati su cui è stata operata la trasformazione. Applicando la trasformata wavelet inversa ai coefficienti così modificati si ottiene una stima di maggiore regolarità del log-spettro della serie. Si considera quindi l’esponenziale di questa stima del log-spettro come una stima dello spettro stesso.

Un approccio molto simile a quello descritto è stato proposto in [24] e [25]. In questi lavori il *wavelet shrinkage* viene effettuato sul log-periodogramma, che costituisce una stima del log-spettro. Le soglie utilizzate sono calcolate *ad hoc* per i vari livelli, tenendo conto dell’approssimazione di Wahba, [46] che scompone il log-periodogramma nella somma del log-spettro vero ed una serie di rumore composto di variabili aleatorie indipendenti ed identicamente distribuite secondo una distribuzione ben specificata. La soglia universale invece produce stime con desiderabili proprietà asintotiche al crescere di n solo nell’ipotesi di dati non correlati e gaussiani, [19].

Dunque anche nella metodologia di stima spettrale proposta in questo lavoro la scelta della soglia universale dovrebbe risultare non ottimale. Tuttavia i risultati ottenuti sono piuttosto soddisfacenti (si veda il riquadro inferiore della figura 3.12) ad eccezione delle regioni dello spettro caratterizzate dalle frequenze di risonanza. Perciò si è adottato una stima mista, con intervalli di 10 Hz attorno alle frequenze di risonanza lasciati senza modifiche dal *wavelet shrinkage*.

È chiaro comunque che questa nuova metodologia di stima spettrale merita un approfondimento della ricerca per definirne le proprietà asintotiche in termini di errore quadratico medio e di *bias*.

4. Si è operata la trasformazione di “whitening” 2.50 sulla serie di 2^{16} dati usando lo spettro stimato sulla serie di rumore correlato lunga 2^{21} dati. La stessa operazione è stata effettuata sui 78 segnali: si noti che in questo

caso il risultato è diverso per ogni simulazione, dal momento che ad ogni simulazione si ha una diversa stima dello spettro di rumore.

5. Con segnale e rumore “sbiancati” sono state effettuate le stesse operazioni impiegate nelle simulazioni con rumore bianco trattate nelle sezioni precedenti.

L'accordo trovato tra simulazioni e calcoli teorici è eccellente. Questo risultato implica che sono trascurabili gli effetti dell'impiego di una ragionevole stima spettrale al posto dello spettro vero $S_f^{(X)}$ nella trasformazione di “whitening” 2.50. Per la statistica $T_{\psi_3,256}$ in particolare l'accordo trovato è ancora migliore rispetto al caso di rumore bianco, come indicato dai valori di ρ mostrati in tabella 3.9: ρ , come in tabella 3.2, indica il rapporto tra le distanze soglia trovate per via teorica e per mezzo di simulazioni con una stessa statistica test.

Questo risultato si giustifica osservando che il rumore alle frequenze più elevate è maggiore rispetto al caso di rumore bianco e che quindi nel processo di “whitening” il segnale viene attenuato in questa regione spettrale in misura maggiore che alle medie frequenze: ne consegue che l'assunzione, richiesta nella trattazione teorica, di un contributo trascurabile del segnale nei coefficienti wavelet alle prime due scale è più facilmente soddisfatta.

Naturalmente le distanze soglia sono state ricalcolate anche per il test del rapporto di verosimiglianza nel caso di rumore correlato. Per questo test si è considerato soltanto il risultato teorico, considerando noto lo spettro di rumore $S_f^{(X)}$: infatti, dal momento che questo test è in ogni caso non applicabile ed il suo ruolo è quello di fornire un test ideale di riferimento, è lecito assumere una condizione ideale anche per quanto riguarda la conoscenza del rumore.

Si è trovato una generale diminuzione delle mediane delle distanze soglia; per il test di massima verosimiglianza si ha $\hat{d}_{\min} = 0.6$ Kpc, $\bar{d} = 16.4$ Kpc, $\hat{d}_{\max} = 73.5$ Kpc. I risultati ottenuti per le statistiche wavelet sono invece riportati in tabella 3.9, mentre la figura 3.13 mostra gli istogrammi delle distanze soglia per alcune delle statistiche considerate.

Era lecito attendersi a priori questa diminuzione delle distanze soglia, rispetto al caso di rumore bianco considerato in precedenza. Infatti, come si può notare nella figura 3.12, la densità spettrale relativa al rumore correlato è più grande, per la maggior parte delle frequenze, rispetto a quella del rumore bianco utilizzato in precedenza. Perciò il rumore che si sta considerando ha

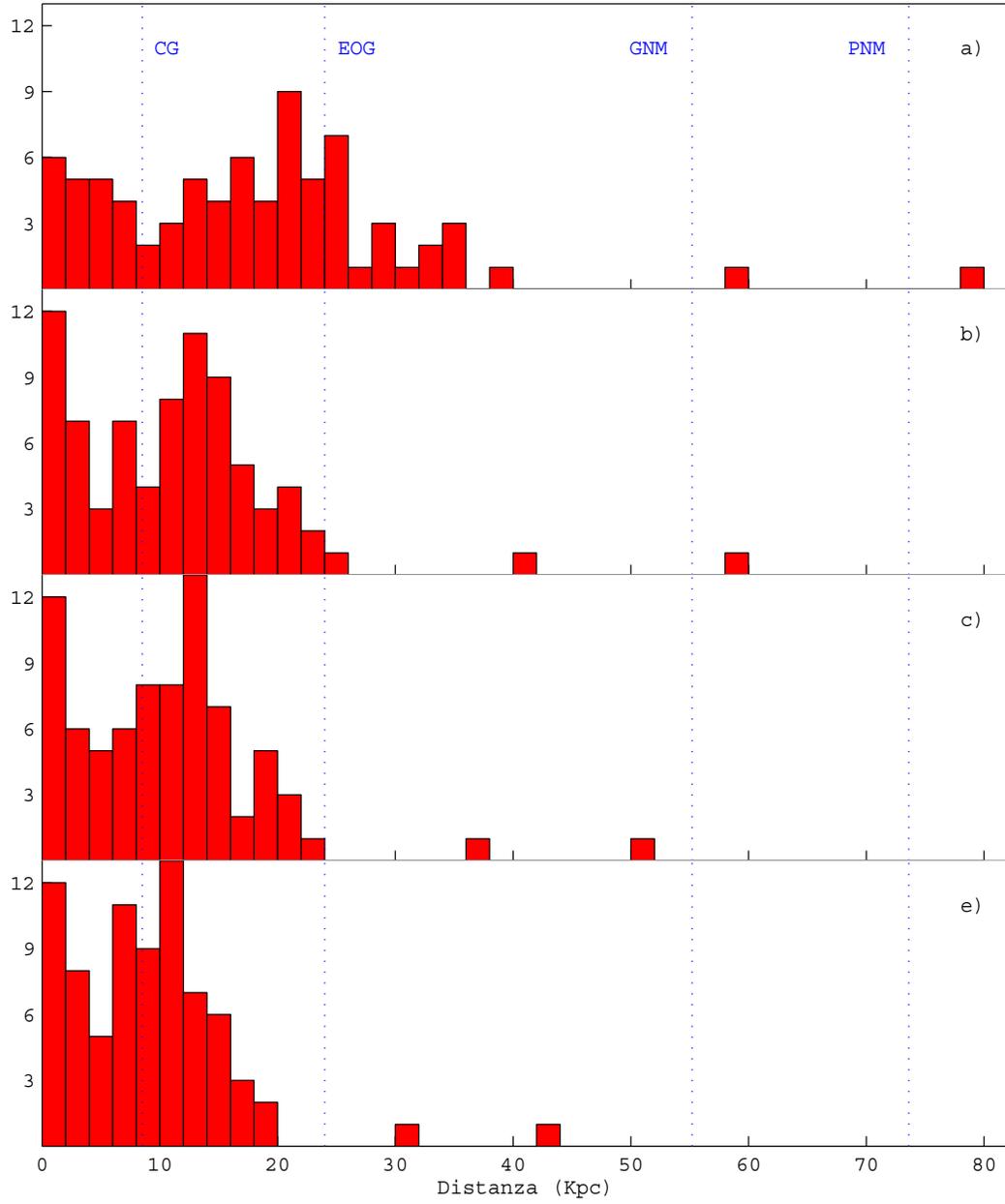


Figura 3.13: Istogrammi delle distanze soglia per i segnali in \mathcal{S}^* in presenza di un rumore correlato con una densità di potenza spettrale come quella mostrata in figura 3.12. Le statistiche test considerate sono: a) rapporto di verosimiglianza; b) $T_{\phi_4,256}$; c) $T_{\psi_4\phi_4,256}$; e) $T_{\psi_3\psi_4\phi_4,256}$. Per le sigle delle linee tratteggiate si faccia riferimento alla figura 3.6.

	$T_{\psi_3,256}$	$T_{\psi_4,256}$	$T_{\phi_4,256}$	$T_{\psi_3\psi_4,256}$	$T_{\psi_4\phi_4,256}$	$T_{\psi_3\psi_4\phi_4,256}$
ν_1	32	16	16	48	32	64
ν_2	192	192	192	192	192	192
$\tau_{\kappa_{\text{fa}}}$	0.4495	0.2923	0.2923	0.5969	0.4495	0.7397
ρ_{\min}	0.9792	0.9701	0.9727	0.9753	0.9764	0.9806
$\bar{\rho}$	1.0046	0.9880	0.9906	0.9899	0.9893	0.9903
ρ_{\max}	1.1436	1.0035	1.0035	1.0052	1.0064	1.0028
σ_ρ	0.0235	0.0075	0.0067	0.0063	0.0060	0.0056
Distribuzioni teoriche						
\hat{d}_{\min}	0.0256	0.2169	0.3912	0.1708	0.4093	0.3472
\bar{d}	1.0406	3.9127	11.4402	3.2206	10.7576	9.0698
\hat{d}_{\max}	3.6951	13.4156	59.1108	10.6499	51.0967	42.7680
Distribuzioni empiriche						
\hat{d}_{\min}	0.0260	0.2206	0.3974	0.1732	0.4167	0.3532
\bar{d}	1.0350	3.9599	11.5668	3.2519	10.8812	9.1622
\hat{d}_{\max}	3.6700	13.5662	60.4759	10.7191	52.1892	43.4792
Distribuzioni teoriche						
ε_{\min}	0.0016	0.0135	0.2171	0.0106	0.4158	0.3617
$\bar{\varepsilon}$	0.0799	0.2770	0.6017	0.2295	0.5966	0.5058
ε_{\max}	0.2138	0.6246	0.7368	0.5111	0.6369	0.5335
σ_ε	0.0558	0.1752	0.1139	0.1430	0.0448	0.0372
Distribuzioni empiriche						
ε_{\min}	0.0016	0.0137	0.2180	0.0108	0.4236	0.3670
$\bar{\varepsilon}$	0.0796	0.2803	0.6075	0.2318	0.6030	0.5107
ε_{\max}	0.2129	0.6370	0.7538	0.5164	0.6505	0.5419
σ_ε	0.0558	0.1776	0.1152	0.1447	0.0452	0.0375

Tabella 3.9: Risultati analoghi a quelli riportati nelle tabelle 3.1, 3.2, 3.4 e 3.5, ma con rumore correlato con densità di potenza spettrale come quella mostrata in figura 3.12. Gli altri parametri sono lasciati invariati. Le distanze sono espresse in kiloparsec. Si noti che i risultati relativi al segnale #16 non sono stati esclusi in nessun caso.

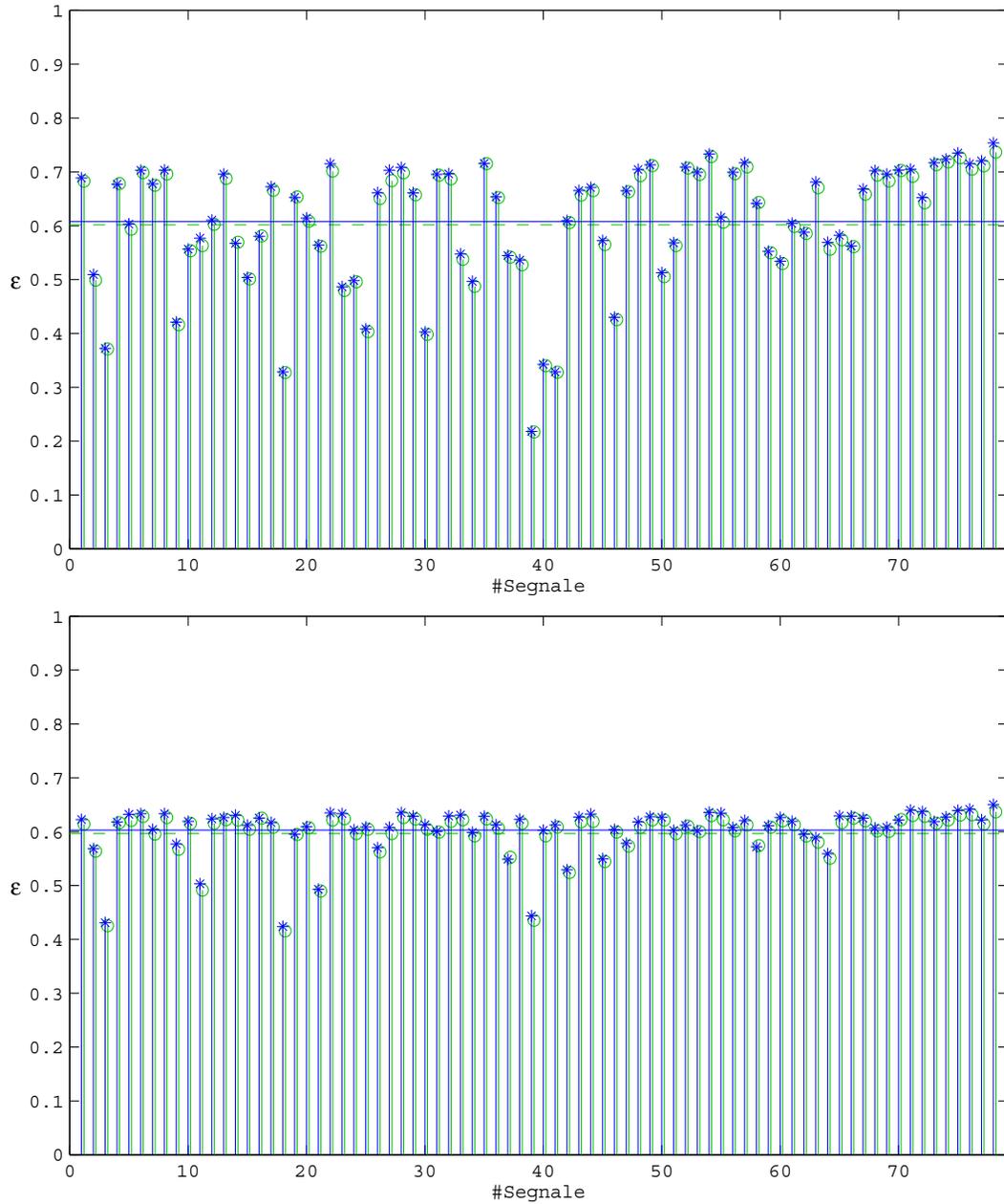


Figura 3.14: Rapporti di prestazioni $\epsilon_{i,T}$, sui 78 segnali di \mathcal{S}^* , ottenuti con test che utilizzano le statistiche $T = T_{\phi_4, 256}$ (in alto) e $T = T_{\psi_4 \phi_4, 256}$ (in basso). I parametri utilizzati sono gli stessi della figura 3.7, ad eccezione del rumore che in questo caso è correlato con densità di potenza spettrale come quella mostrata in figura 3.12.

un'energia complessiva molto maggiore di quello utilizzato nella sezione 3.2. La differenza tra i due spettri è di molti ordini di grandezza, a parte che nella regione delle medie frequenze, $10^2 - 10^3$ Hz, dove risultano comparabili. D'altra parte è proprio questa la banda di frequenze di maggior rilievo per i segnali che si intende rivelare.

Dal confronto delle tabelle 3.9 e 3.5 si osserva che per tutte le statistiche wavelet, con l'eccezione della $T_{\phi_4,256}$, il valore di $\bar{\varepsilon}$ è maggiore rispetto al caso di rumore bianco. In effetti, con uno spettro dall'andamento concavo come quello mostrato in figura 3.12, la trasformazione 2.50 attenua il contributo dei segnali alle basse ed alte frequenze. La statistica $T_{\phi_4,256}$, che ha la struttura di un filtro passa-basso, risente più delle altre dell'attenuazione del segnale alle basse frequenze: si spiega così la diminuzione di $\bar{\varepsilon}$ rispetto al caso di rumore bianco considerato in precedenza. Viceversa le altre statistiche wavelet risultano particolarmente sensibili anche ai contributi del segnale alle medie frequenze, quelle meno attenuate dal processo di *whitening*.

In particolare si nota che la statistica $T_{\psi_4\phi_4,256}$ nel caso di rumore correlato ottiene praticamente lo stesso valore di $\bar{\varepsilon}$ conseguito dalla $T_{\phi_4,256}$, garantendo nel contempo una robustezza assai maggiore rispetto allo specifico segnale di \mathcal{S}^* : infatti mostra un valore sensibilmente inferiore per σ_ε . Le proprietà di robustezza della $T_{\psi_4\phi_4,256}$ sono evidenziate in figura 3.14, dove si può notare la grande omogeneità dei valori del rapporto di prestazioni $\varepsilon_{i,T}$ sui 78 segnali di \mathcal{S}^* .

3.4 Densità di probabilità di rivelazione

Nel discutere le proprietà delle varie statistiche test proposte per il problema della rivelazione di onde gravitazionali prodotte in una esplosione di supernova, fino a questo punto non si è tenuto conto della distribuzione spaziale delle possibili sorgenti. In realtà sarebbe opportuno considerare anche questa informazione nel confrontare i risultati di metodologie di rivelazione diverse. Infatti non è lecito attendersi una medesima probabilità a priori per tutti i valori della distanza d della sorgente: da un lato il volume di spazio esaminato cresce con la terza potenza di d , dall'altro le stelle progenitrici di supernovae non sono omogeneamente distribuite nello spazio, almeno per l'ordine di

distanze considerato in questo lavoro.

All'atto pratico due test con distanze soglia rispettivamente di 35 e 45 Kpc hanno la stessa utilità per l'applicazione considerata: infatti entrambi riuscirebbero agevolmente a rivelare un'eventuale supernova all'interno della nostra galassia, ma nessuno dei due ha una potenza adeguata alla rivelazione di supernovae in altre galassie, incluse le nubi di Magellano che sono i sistemi stellari piú prossimi al nostro.

Nel corso del presente lavoro sono state derivate, per alcuni test, le probabilità di rivelazione dell' i -esimo segnale in \mathcal{S}^* condizionate alla distanza d della sorgente. Dalla proposizione 2.5 si deduce che, per una generica statistica T tra quelle proposte nella sezione 2.2, tale probabilità condizionata vale:

$$\Pr [i | d] = F_{\nu_1, \nu_2, \delta}^{(CD)} \left(\frac{\nu_2}{\nu_1} \tau_{\kappa_{\text{fa}}} \right), \quad (3.16)$$

dove $F_{\nu_1, \nu_2, \delta}^{(CD)}$ indica una distribuzione di Fisher cumulata da destra con ν_1 e ν_2 gradi di libertà e parametro di non centralità δ . Si ricorda che i gradi di libertà sono determinati dal numero di coefficienti sommati a numeratore ed a denominatore nella statistica T , mentre il parametro di non centralità dipende dalla distanza d e dalla forma d'onda del segnale considerato, sottoposta all'algoritmo di *whitening* utilizzato per rimuovere la correlazione del rumore. Il valore di soglia $\tau_{\kappa_{\text{fa}}}$ è fissato dal tasso di falsi allarmi che si è disposti a tollerare.

Sfruttando la proprietà 3.1 ed il corollario 3.2 si ricava un'analoga probabilità condizionata anche per il test del rapporto di verosimiglianza:

$$\Pr [i | d] = N_{\|\tilde{s}\|, 1}^{(CD)} \left(\Phi_{(1-\alpha)} \right), \quad (3.17)$$

dove $N_{\|\tilde{s}\|, 1}^{(CD)}$ è la cumulata da destra di una distribuzione normale di varianza unitaria e media data dalla norma euclidea di \tilde{s} . Questo segnale è ottenuto dalla trasformazione $\infty 26s = \Gamma s$, nella quale s è il segnale gravitazionale normalizzato per la distanza d e Γ è una matrice tale che $\Gamma' \Gamma = \Sigma^{-1}$ (con Σ matrice di varianza e covarianza del rumore). $\Phi_{(1-\alpha)}$ è il quantile per la probabilità $(1 - \alpha)$ di una distribuzione normale standard; α è l'ampiezza del test, fissata sulla base del tasso di falsi allarmi κ_{fa} che si è disposti a tollerare.

D'altra parte conoscendo la densità di probabilità a priori $p[d]$ che un evento di supernova si verifichi alla distanza d dalla Terra, potrebbe essere

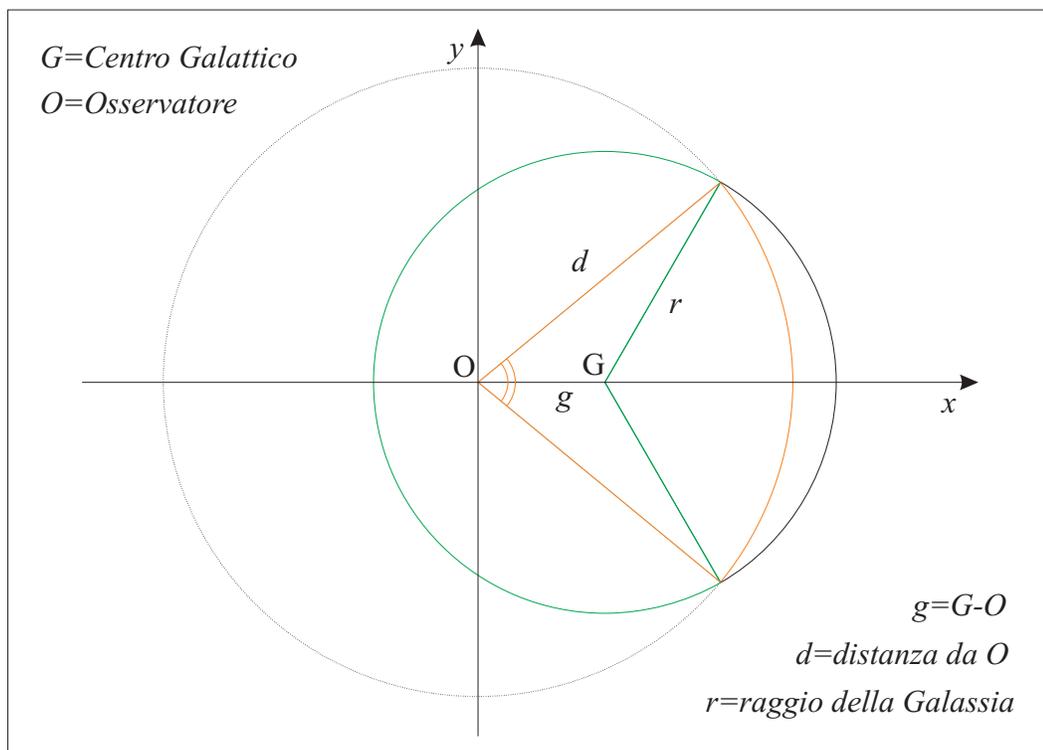


Figura 3.15: Modello schematico della geometria del disco galattico.

interessante considerare la densità probabilità per la rivelazione dell' i -esimo segnale di \mathcal{S}^* alla distanza d :

$$p[i \& d] = \Pr[i | d] p[d]. \quad (3.18)$$

Questa relazione permette dunque di valutare la combinazione delle caratteristiche di rivelazione della statistica test con le esigenze di rivelazione poste dal problema applicativo.

Il problema da risolvere a questo punto sarebbe quello di specificare una ragionevole densità probabilità a priori $p[d]$ per la distribuzione della distanza d di un evento di supernova. È noto che tali eventi avvengono di solito all'interno di nubi di idrogeno ionizzato, nubi che nelle galassie a spirale come la nostra si dispongono prevalentemente su un disco sottile. Dagli studi di astrofisica sulle nubi di idrogeno ionizzato nella nostra galassia si potrebbe pertanto ottenere una ragionevole specificazione della densità $p[d]$, ma si tratta in ogni caso di un compito che va al di là degli scopi del presente lavoro.

A puro titolo di esempio si è fatto un modello estremamente semplificato, restringendo l'attenzione alla nostra galassia schematizzata come un disco sottile di raggio $r = 15.5$ Kpc ed assumendo che la probabilità di osservare una supernova entro una data area all'interno del disco galattico sia proporzionale all'area stessa. In questo caso, ricordando che la Terra si trova ad una distanza di $g = 8.5$ Kpc dal centro galattico si ottiene:

$$p[d] = \begin{cases} 2\pi d, & \text{se } d \leq r - g; \\ 2d \arccos\left(\frac{d^2 + g^2 - r^2}{2gd}\right), & \text{se } d \in (r - g; r + g); \\ 0, & \text{altrimenti;} \end{cases} \quad (3.19)$$

che corrisponde, nella figura 3.15, alla lunghezza dell'arco di circonferenza di raggio d e centro in O che interseca il cerchio di raggio r con centro in G. In un modello più realistico d'altra parte si dovrebbe considerare una probabilità di eventi per unità di superficie in funzione della distanza dal centro della galassia anziché costante.

Ad ogni modo nella figura 3.16 vengono mostrate, in funzione di d , le densità di probabilità, $p[i \& d]$ e $p[d]$ insieme alla probabilità condizionata $\Pr[i | d]$ per la statistica $T_{\phi_4, 256}$ applicata ai segnali #72 e #1; nella stessa figura vengono riportati anche i risultati corrispondenti ottenibili con il test del rapporto di verosimiglianza.

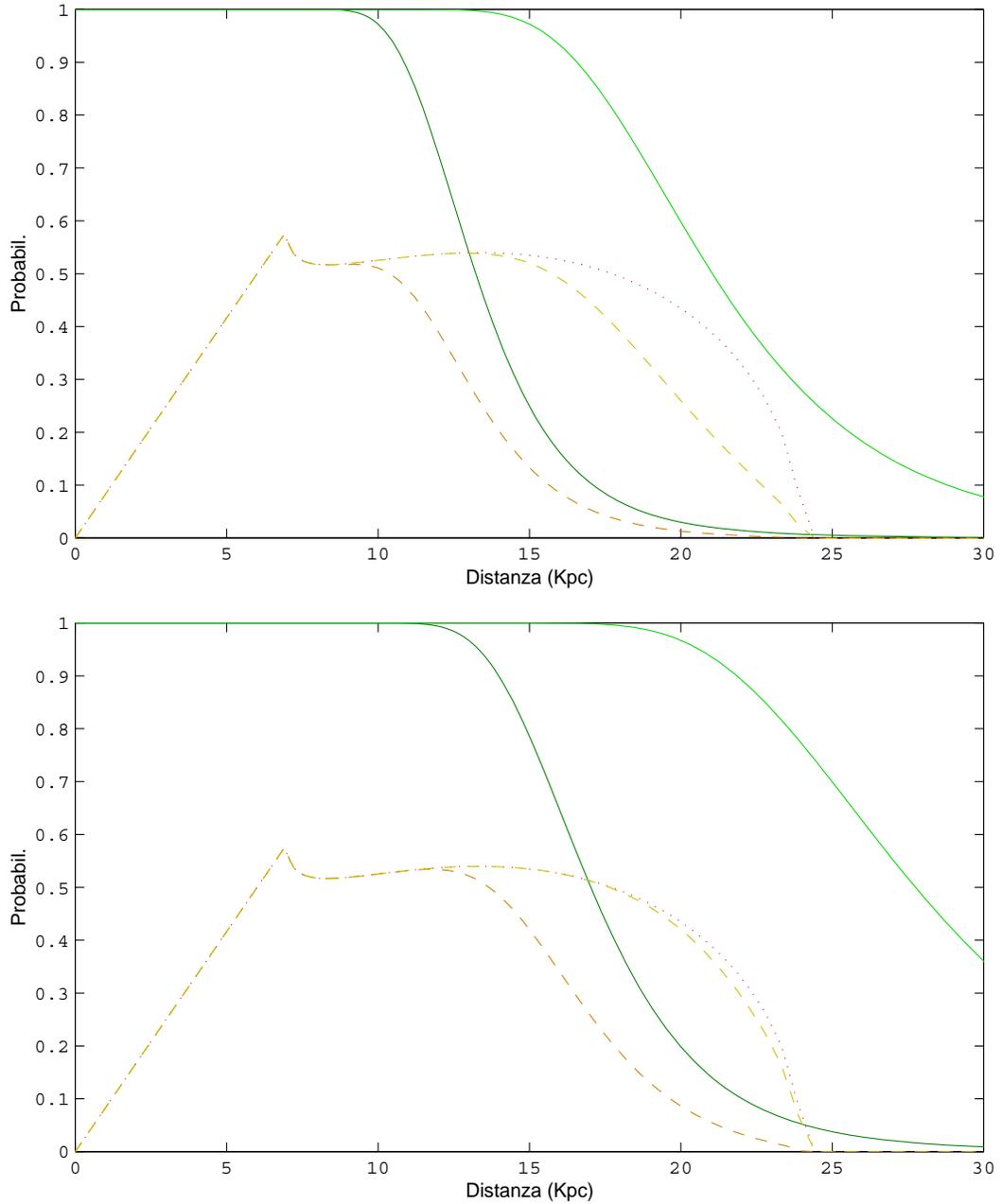


Figura 3.16: *In alto*: probabilità condizionata $\Pr[i|d]$ per la rivelazione del segnale $i = \#72$ con la statistica $T_{\phi_4, 256}$ (linea continua più scura) e con il rapporto di verosimiglianza (linea continua più chiara) e densità di probabilità di rivelazione $p[i \& d]$ corrispondenti (linee tratteggiate). La densità di probabilità a priori utilizzata $p[d]$ è mostrata con linea punteggiata. $p[d]$ e le $p[i \& d]$ sono ingrandite di un fattore 10. *In basso*: grafico analogo al precedente, ma per il segnale $i = \#1$.

Capitolo 4

Conclusioni

4.1 Discussione

In questo lavoro è stato affrontato il problema dell'individuazione di un segnale transiente in una serie storica di rumore gaussiano e almeno localmente stazionario, disponendo di una conoscenza assai vaga sulle caratteristiche del segnale da rivelare, limitata a grossolane indicazioni sulla sua durata ed ampiezza di banda spettrale. Si è supposto inoltre di dover trattare una considerevole mole di dati, tale da rendere di primaria importanza le considerazioni sulla rapidità computazionale degli algoritmi proposti per la soluzione del problema descritto. Non sono invece stati posti vincoli sulla struttura di correlazione del rumore, a patto che la si potesse considerare stazionaria per tempi grandi rispetto alla durata del segnale transiente. Infatti, pur partendo dal caso più elementare di rumore bianco, sono state discusse e verificate le opportune modifiche per estendere la trattazione al caso di rumore correlato.

Si è cercato una soluzione a questo tipo di problema rappresentando la serie storica nel dominio tempo-frequenza, per poter sfruttare le conoscenze a priori sull'ordine di grandezza della durata e della larghezza di banda dei segnali da rivelare. Per questo scopo lo strumento matematico delle wavelets appare di particolare interesse grazie alla proprietà di una buona localizzazione temporale di picchi di breve durata affiancata da una discreta risoluzione alle basse frequenze, ma grazie anche all'esistenza di algoritmi a basso costo computazionale per il calcolo delle trasformate wavelet.

Sono quindi state definite alcune statistiche test, in parte originali, costruite sulla base dei coefficienti di una trasformata wavelet discreta, ortogonale e a

supporto compatto, della serie storica sotto esame. Questo tipo di analisi a più livelli di risoluzione offre l'opportunità di ridurre al minimo sia il numero di coefficienti da considerare, grazie all'ortogonalità, sia il costo computazionale necessario per calcolare tali coefficienti, grazie al supporto compatto. La wavelet madre è stata scelta come compromesso tra l'esigenza di un andamento non troppo irregolare e quella di una buona capacità di localizzazione temporale. I parametri delle statistiche test proposte devono essere scelti in modo da trarre vantaggio dalle poche conoscenze sul segnale discusse in precedenza.

I test di ipotesi vengono quindi implementati confrontando il valore ottenuto dalle statistiche con un valore di soglia fissato sulla base del massimo tasso di falsi allarmi che si è disposti a tollerare nell'esperimento.

Come applicazione concreta del lavoro teorico svolto si è affrontato uno specifico problema con i requisiti definiti poco sopra: la rivelazione con lo strumento VIRGO di segnali gravitazionali emessi in eventi di supernova, con collasso del nucleo centrale a stella di neutroni. I modelli teorici per questo tipo di eventi sono piuttosto incerti e conducono a previsioni sulle caratteristiche del segnale osservabile assai diverse tra loro. Tuttavia si può grossolanamente accettare una durata dei segnali dell'ordine di poche decine di millisecondi ed una densità di potenza spettrale concentrata al di sotto di 1 KHz.

Da un punto di vista fisico i risultati presentati in questo lavoro non sembrano essere molto incoraggianti: al massimo si può sperare di osservare segnali provenienti da una supernova esplosa nella nostra galassia. Si tratta di eventi che si verificano piuttosto raramente, con un valore atteso dell'ordine di uno o due eventi per secolo. D'altra parte questo non è tanto un limite della metodologia proposta, quanto piuttosto dello stato dell'arte della tecnologia utilizzata per la realizzazione dell'esperimento : infatti i modelli teorici, per quanto incerti, allo stato attuale prevedono dei segnali intrinsecamente troppo deboli per essere rivelati a distanze extragalattiche, qualsiasi sia la metodologia adottata. D'altra parte, visto l'alto grado d'incertezza nelle conoscenze dei dettagli della fisica di questi eventi, non si può escludere che la natura sia più benevola di quanto non ci si aspetti. Ad esempio alcuni studi, [29], suggeriscono un'emissione gravitazionale più intensa nelle fasi successive all'esplosione di una supernova, fasi per le quali i modelli considerati in questo lavoro rinunciavano a fornire qualsiasi tipo di previsione.

Inoltre l'esperimento VIRGO può contare su altri eventi astrofisici in grado di produrre segnali gravitazionali più intensi: si deve infatti sottolineare che l'individuazione anche di un solo segnale di questo tipo sarebbe un successo di fondamentale importanza, dal momento che sino ad ora si sono avute soltanto conferme indirette della teoria delle onde gravitazionali. Tra gli eventi più promettenti deve essere menzionato il caso dell'emissione gravitazionale da parte di binarie coalescenti¹. Per questa classe di eventi le teorie sembrano essere piuttosto affidabili e forniscono indicazioni assai precise sulla forma del segnale da ricercare: perciò sarà possibile utilizzare un test ottimale fondato sul rapporto di verosimiglianza, chiamato in questo contesto *matched filtering*. Ci preme tuttavia sottolineare che, come alternativa al *matched filtering*, anche per la rivelazione di questo tipo di segnali è stata proposta una metodologia fondata sull'analisi in wavelet: in [30] si propone di cercare il massimo dell'integrale di linea in una classe di curve definite nel dominio della trasformata wavelet continua del segnale, per poi confrontare questo valore massimo con un valore di soglia stabilito in funzione del tasso di falsi allarmi che si è disposti a tollerare. Tra i possibili elementi a favore di questa metodologia potrebbero esserci una migliore efficienza computazionale ed una maggiore robustezza, in particolare nel caso, per la verità assai improbabile, di due eventi quasi contemporanei.

Tornando al presente lavoro deve essere sottolineato che i risultati ottenuti, per quanto non molto incoraggianti dal punto di vista fisico, debbono essere considerati estremamente soddisfacenti dal punto di vista statistico. Infatti la metodologia proposta è ben comparabile con le migliori altre presentate in letteratura come possibili soluzioni per lo stesso problema, manifestando nel contempo evidenti indicazioni di una maggiore robustezza rispetto alle possibili forme del segnale ignoto da rivelare. Le prestazioni ottenute sono state valutate per mezzo di un confronto con i risultati ottenibili con il test del rapporto di verosimiglianza, metodologia ottimale ma applicabile solo idealmente poiché richiede la conoscenza del segnale da rivelare. Si è trovato che la distanza della sorgente per cui si ha il 50 % di probabilità di rivelazione con la metodologia proposta risulta con una certa stabilità dell'ordine di 0.6 volte quella ottenibile con la metodologia ottimale. Una tale diminuzione implica

¹Si tratta di oggetti condensati, come stelle di neutroni o buchi neri, che orbitano l'uno intorno all'altro avvicinandosi sempre più, fino a fondersi in un unico oggetto.

una limitazione piuttosto rilevante del volume di spazio esplorabile e, tenendo conto della distribuzione spaziale delle possibili sorgenti, comporta un sensibile calo della probabilità di rivelare un evento entro un fissato intervallo di tempo. Ad ogni modo non si ottengono risultati migliori con le altre metodologie effettivamente applicabili.

Inoltre l'originale classe di test wavelet proposta in questo lavoro manifesta doti di robustezza rispetto alla forma dei segnali transienti che non sono condivise da altri test di ipotesi, di potenza comparabile, proposti in letteratura per la soluzione del medesimo problema.

Ricordando che i segnali gravitazionali hanno un'ampiezza inversamente proporzionale alla distanza della sorgente, si può dire che con la metodologia introdotta in questo lavoro vengono rivelati segnali con un rapporto segnale-rumore maggiore di un fattore $1/0.6 \simeq 1.7$ rispetto a quello dei segnali rivelabili idealmente dalla metodologia ottimale. La formulazione dei risultati ottenuti in termini di ampiezza piuttosto che di distanza risulta di maggiore generalità, allargando l'interesse anche a segnali di natura non astrofisica, come ad esempio alcune forme di rumore strumentale di breve durata. È fondamentale infatti sottolineare la generalità della metodologia proposta, che non si adatta specificamente al particolare problema applicativo che si è scelto di affrontare.

Al di là dell'esperimento VIRGO, la metodologia wavelet discussa in questo lavoro risulta indicata per l'applicazione in qualsiasi altro contesto nel quale si richieda la rivelazione di eventi transienti in una serie storica caratterizzata da un rumore gaussiano localmente stazionario ed additivo. In particolare questa tecnica risulta estremamente raccomandabile per serie storiche con una grande mole di dati, per le quali possano rappresentare un problema i costi computazionali richiesti da altri tipi di analisi.

4.2 Problemi aperti e sviluppi futuri

I risultati conseguiti incoraggiano a continuare ad utilizzare gli strumenti offerti dalla teoria delle wavelets nell'analisi di serie storiche volta alla ricerca di segnali nascosti da rumore. È piuttosto ampio il ventaglio di scelte possibili per continuare a sviluppare metodologie di analisi con le stesse premesse e fina-

lità discusse in questo lavoro: dalla trasformata wavelet continua ai pacchetti wavelet introdotti in [11].

In particolare desta un certo interesse la possibilità di caratterizzare anche la componente localmente stazionaria del rumore sfruttando rappresentazioni tempo-frequenza come quelle ottenute con le trasformate wavelet. Lavori interessanti in questa direzione sono stati presentati in [36] e [20], ma senza dubbio si tratta di una linea di ricerca ancora non adeguatamente esplorata e, proprio per questo, potenzialmente molto promettente.

Ad ogni modo restano da chiarire anche alcuni aspetti del lavoro svolto e presentato in questa sede. In primo luogo è auspicabile un approfondimento nello studio della dipendenza dei risultati dalle scelte operate sui vari parametri che concorrono nella definizione delle statistiche test: in particolare dovrebbero essere verificate sia l'influenza della scelta della wavelet madre ψ , sia la dipendenza dalla separazione tra intervalli di analisi consecutivi.

L'effetto di diverse scelte per l'ampiezza del singolo intervallo è stato discusso anche nel presente lavoro ed è risultato piuttosto rilevante. Viste le assunzioni fatte sulla conoscenza a priori dei segnali da rivelare, resta un problema piuttosto critico scegliere un valore unico n per la larghezza degli intervalli che sia adeguato e soddisfacente per tutti i possibili segnali. Perciò sarebbe di grande importanza riuscire a modificare le statistiche test in modo che risultino assai meno sensibili alla scelta operata per il parametro n . Si è pensato che questo risultato potrebbe essere ottenuto limitando la sommatoria nei numeratori delle statistiche 2.11 e 2.24 soltanto ai termini di maggiore entità, tralasciando la somma dei valori più piccoli di una certa soglia. In questo modo il numero di termini di tali sommatorie non sarebbe direttamente determinato dal valore n che si è fissato, ma in qualche misura si adatterebbe alle caratteristiche del segnale eventualmente presente nell'intervallo di dati analizzato. L'idea pare molto interessante, purtroppo però si scontra con la difficoltà di ricavare per le statistiche test così modificate le distribuzioni di probabilità teoriche, sia sotto ipotesi alternativa che sotto ipotesi nulla. D'altra parte il calcolo di tali distribuzioni per mezzo di simulazioni appare piuttosto lungo ed impegnativo, dal momento che si è interessati spesso alle code di tali distribuzioni, a causa dei valori estremamente ridotti del massimo tasso di falsi allarmi che si è in genere disposti a tollerare.

Un altro problema rimasto per adesso senza una soluzione riguarda lo studio della correlazione dei valori delle statistiche test calcolate in diversi intervalli di dati parzialmente sovrapposti. Come evidenziato nella sezione 2.2.2, non è del tutto soddisfacente fissare le soglie dei test sulla base del valore atteso del massimo tasso di falsi allarmi tollerabile: infatti lo stesso falso allarme verrà segnalato in una serie di intervalli consecutivi, dal momento che le statistiche test in tali intervalli sono tra loro correlate. Sarebbe dunque più idoneo fissare la probabilità che in un determinato periodo di tempo si osservi almeno un falso allarme. Nella sezione 2.2.2 si è provato a calcolare questa probabilità seguendo un parallelo con i risultati trovati per la *Scan Statistic*, arrestandosi però di fronte alle differenze tra la natura discreta di tale statistica e quella continua delle statistiche test discusse nel presente lavoro. Questo tipo di calcolo resta dunque forse il principale problema da risolvere nel contesto della metodologia discussa nel presente lavoro.

Infine, per quanto riguarda lo specifico problema della rivelazione di segnali gravitazionali generati in esplosioni di supernovae, resta da definire un adeguato modello per la distribuzione nello spazio delle possibili sorgenti di segnale. Anche se questa è in effetti una richiesta da porre alla comunità dei fisici, deve esserne sottolineata la profonda rilevanza per qualsiasi metodologia statistica si intenda utilizzare per la soluzione del problema applicativo discusso in questo lavoro. Si tratta in effetti di una ulteriore informazione a priori che potrebbe risultare assai preziosa nella costruzione di nuove statistiche test, oltre che per un confronto più realistico dei risultati effettivamente ottenibili con le varie statistiche già proposte.

Appendice A

Le Wavelets

Questa appendice costituisce un'introduzione alla teoria delle wavelets molto elementare ed ha lo scopo di fornire le nozioni di base necessarie per comprendere alcuni aspetti delle metodologie introdotte e discusse nel presente lavoro. La letteratura disponibile per approfondire gli argomenti trattati è assai vasta, sia in termini di articoli pubblicati su riviste che di monografie. Per iniziare ad orientarsi in tale abbondanza di trattazioni, si segnalano testi classici come [10], [15], [51] o più recenti come [22], [44] e [39].

A.1 Trasformate Wavelet

Data una funzione di base $\psi \in L_2(\mathbb{R})$ che soddisfi la seguente *condizione di ammissibilità*:

$$C_\psi = \int_{\mathbb{R}} \frac{|F^{(\psi)}(\omega)|^2}{\omega} d\omega < +\infty, \quad (\text{A.1})$$

dove $F^{(\psi)}(\omega)$ è la trasformata di Fourier di ψ , la *trasformata wavelet continua* di una funzione $f(x)$ rispetto alla *wavelet madre* ψ è definita come:

$$W_\psi^{(f)}(a, b) = \int_{\mathbb{R}} f(x) \overline{\psi_{a,b}(x)} dx, \quad (\text{A.2})$$

In questa espressione le funzioni:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right), \quad (\text{A.3})$$

sono *trasformazioni affini* di ψ , cioè sono ottenute a partire da ψ con una dilatazione di un fattore a , detto *parametro di scala*, ed una traslazione di

una quantità b , detta *parametro di traslazione*. La costante di normalizzazione $1/\sqrt{|a|}$ è scelta in modo da garantire la conservazione della norma in $L_2(\mathbb{R})$ per tutte le wavelets definite dalla A.3:

$$\int_{\mathbb{R}} |\psi_{a,b}(x)|^2 dx = \int_{\mathbb{R}} |\psi(x)|^2 dx, \quad \forall a. \quad (\text{A.4})$$

Esiste anche una *trasformata wavelet continua inversa* grazie alla quale è possibile ricostruire la funzione $f(x)$ a partire dalla sua trasformata $W_{\psi}^{(f)}(a, b)$:

$$f(x) = \frac{1}{C_{\psi}} \int_{\mathbb{R}^2} W_{\psi}^{(f)}(a, b) \psi_{a,b}(x) \frac{da}{a^2} db. \quad (\text{A.5})$$

Il denominatore a^2 è conseguenza della scelta fatta per la normalizzazione delle funzioni $\psi_{a,b}(x)$ definite dalla relazione A.3. In letteratura si trovano anche scelte diverse per questa normalizzazione e ciò comporta una diversa formulazione della trasformata inversa data dalla A.5. Ad esempio in [12] si propone l'utilizzo di un fattore di normalizzazione pari ad a^{-1} grazie al quale viene conservata la norma delle wavelets $\psi_{a,b}(x)$ in $L_1(\mathbb{R})$ anziché in $L_2(\mathbb{R})$: con questa scelta nell'equazione A.5 il termine da/a^2 viene sostituito da $da/a = d(\ln(a))$.

Da quanto esposto fino a questo punto si nota una certa analogia della trasformata wavelet con la trasformata di Fourier: entrambe possono essere poste nella forma:

$$T^{(f)}(p) = \int_{\mathbb{R}} f(x) \overline{k_p(x)} dx, \quad (\text{A.6})$$

con $p = \omega = 2\pi\nu$ e $k_p(x) = e^{i\omega x}$ nel caso della trasformata di Fourier e $p = (a, b)$ e $k_p(x) = \psi_{a,b}(x)$ per la trasformata wavelet. In entrambi i casi è disponibile una trasformata inversa in cui $k_p(x)$ compare come uno dei fattori della funzione integranda. Ma soprattutto per entrambe le trasformate $k_p(x)$ ha un andamento oscillante, con rapidità delle oscillazioni controllata dal parametro ω o a . Per le wavelets questo tipo di andamento è imposto dalla condizione di ammissibilità A.1 ed è all'origine del loro nome che potrebbe essere tradotto in *ondine*.

Al di là di queste analogie formali ci sono importanti differenze strutturali tra i due tipi di trasformate. Infatti nel caso della trasformata wavelet la condizione di ammissibilità A.1 implica un decadimento a zero della funzione $k_p(x)$ per $|x| \rightarrow +\infty$, e quindi si può dire che $k_p(x)$ è localizzata attorno ad

un valore di x determinato dal parametro b . Per la trasformata di Fourier $k_p(x)$ è ottenuto come combinazione lineare di seni e coseni ed è per questo motivo che non ha alcun parametro che ne caratterizzi la localizzazione lungo la coordinata x .

L'analogia diviene più stretta qualora si confronti la trasformata wavelet con una versione modificata della trasformata di Fourier che va sotto il nome di trasformata di Fourier con finestra, o *Short Time Fourier Transform*, introdotta proprio con lo scopo di ottenere quella proprietà di localizzazione in x di cui non gode la trasformata di Fourier classica. In questo caso nell'equazione A.6 si ha $k_p(x) = g(x - x_0)e^{i\omega x}$ con $g(x)$ *funzione finestra* localizzata attorno a $x_0 + \int_{\mathbb{R}} xg(x)dx$.

Entrambe le trasformate consentono quindi di scomporre la funzione f in una serie di contributi dati da oscillazioni *locali* più o meno rapide. La rapidità delle oscillazioni è controllata dal parametro di scala a nella trasformata wavelet e dalla frequenza angolare ω per la trasformata di Fourier con finestra. Il parametro di scala a può essere considerato come inversamente proporzionale ad una frequenza ω . La localizzazione, nelle due trasformate, è legata invece ai parametri b ed x_0 .

In realtà non viene mai isolato il singolo contributo relativo alla locazione ed alla frequenza specificate dalla coppia di parametri; piuttosto si mescolano i contributi di un intervallo di frequenze ed un intervallo di locazioni attorno a quelle specificate. Questo tipo di inconveniente non è eliminabile contemporaneamente per entrambi i dominî, quello delle locazioni e quello delle frequenze. Infatti si può dimostrare il seguente *principio di indeterminazione*:

$$\Delta x \cdot \Delta \omega \geq \text{cost} \tag{A.7}$$

secondo il quale non è possibile ottenere un'elevata risoluzione in frequenza ($\Delta\omega$ piccolo) se non a scapito della risoluzione nella locazione, cioè con un aumento di Δx . Le rappresentazioni della funzione f nel dominio delle x e in quello di Fourier corrispondono ai due casi estremi di questo principio: la prima ha la massima risoluzione in x e nessuna risoluzione in frequenza, la seconda esattamente il contrario. Le rappresentazioni ottenute con la trasformata wavelet e la trasformata di Fourier con finestra viceversa hanno risoluzione finita e non nulla in entrambi i dominî. La differenza tra queste due trasformate sta nelle modalità con cui viene rispettato il vincolo imposto dalla A.7: per la tra-

sformata di Fourier con finestra Δx e $\Delta\omega$ sono costanti per tutte le locazioni e per tutte le frequenze, mentre per la trasformata wavelet vengono mantenuti costanti $\omega\Delta x$ e $\Delta\omega/\omega$. Ciò vuol dire che nella trasformata wavelet si ha una risoluzione in frequenza migliore per le basse frequenze, mentre per i fenomeni rapidi si privilegia la risoluzione in x : evidentemente questa è una proprietà assai utile in presenza di non stazionarietà locali di breve durata, come ad esempio dei segnali transienti rapidi sovrapposti ad un rumore stazionario di bassa frequenza.

Spesso la coordinata di locazione x è in effetti un tempo, perciò la trasformata di Fourier con finestra e la trasformata wavelet prendono il nome di *rappresentazioni tempo-frequenza lineari*. Esistono anche rappresentazioni quadratiche che decompongono nel dominio tempo-frequenza l'energia della funzione f piuttosto che non direttamente f , ottenendo delle distribuzioni la cui marginale rispetto alla frequenza tende alla densità di potenza spettrale di f . Si può dimostrare che il modulo quadro sia della trasformata wavelet che di quella di Fourier con finestra corrispondono ad opportuni *smoothing* della più nota delle rappresentazioni tempo-frequenza quadratiche, la distribuzione di Wigner-Ville, [41].

A.1.1 La trasformata wavelet discreta

La trasformata wavelet continua risulta strettamente correlata per scale o per locazioni la cui separazione sia inferiore al corrispettivo intervallo di risoluzione. In effetti la rappresentazione fornita dalla trasformata wavelet continua è in qualche modo ridondante. Sotto certe condizioni è possibile definire una formula di ricostruzione della funzione f che utilizzi il valore della sua trasformata wavelet $W_\psi^{(f)}(a, b)$ calcolata soltanto per un insieme numerabile di coppie di parametri (a, b) , ponendo $a = a_0^j$ e $b = b_0 a_0^j m$, con $j, m \in \mathbb{Z}$ e a_0, b_0 fissati.

Un caso particolare di questa situazione si ha qualora sia soddisfatta la condizione:

$$\|f\|^2 \leq \sum_{j,m} \langle f | \psi_{j,m} \rangle^2 \leq \|f\|^2, \quad \forall f \in L_2(\mathbb{R}), \quad (\text{A.8})$$

dove $\langle f|g \rangle = \int_{\mathbb{R}} f(x)\overline{g(x)}dx$ è l'usuale prodotto scalare in $L_2(\mathbb{R})$ e si è posto:

$$\psi_{j,m}(x) = \frac{1}{\sqrt{a_0^j}} \psi \left(\frac{x - b_0 a_0^j m}{a_0^j} \right). \quad (\text{A.9})$$

In questo caso funzioni $\psi_{j,m}(x)$ costituiscono una base ortonormale di $L_2(\mathbb{R})$ ed è pertanto possibile utilizzare la formula:

$$f(x) = \sum_{j,m} \langle f|\psi_{j,m} \rangle \psi_{j,m}(x), \quad (\text{A.10})$$

per ricostruire la funzione $f(x) \in L_2(\mathbb{R})$ a partire dai coefficienti della *trasformata wavelet discreta* $c_{j,m}^{(f)} = \langle f|\psi_{j,m} \rangle$. Nel caso che si incontra più comunemente la trasformata utilizza i valori $a_0 = 1/2$ e $b_0 = 1$ e d'ora in avanti assumeremo implicitamente questa scelta.

La trattazione della scomposizione di una funzione $f \in L_2(\mathbb{R})$ su una base wavelet ortonormale può essere formalizzata nel contesto della teoria dell'*analisi a più livelli di risoluzione*. Formalmente si tratta di una successione V_r , $r \in \mathbb{Z}$, di sottospazi chiusi di $L_2(\mathbb{R})$ che soddisfano le seguenti condizioni:

$$V_r \subset V_{r+1}; \quad (\text{A.11})$$

$$\overline{\bigcup_{r \in \mathbb{Z}} V_r} = L_2(\mathbb{R}); \quad (\text{A.12})$$

$$\bigcap_{r \in \mathbb{Z}} V_r = \emptyset; \quad (\text{A.13})$$

$$f(x) \in V_r \Leftrightarrow f(2x) \in V_{r+1}; \quad (\text{A.14})$$

$$f(x) \in V_r \Leftrightarrow f(x+m) \in V_r, \forall m \in \mathbb{Z}; \quad (\text{A.15})$$

$$\exists \phi \in V_0 : \{\phi(x-m), m \in \mathbb{Z}\} \text{ è una base ortonormale in } V_0. \quad (\text{A.16})$$

La funzione $\phi(x)$ è detta funzione di scala. Applicando iterativamente la A.14 si osserva che:

$$f(x) \in V_0 \Leftrightarrow (2^r x) \in V_r, \quad (\text{A.17})$$

e tenendo conto anche della A.15 e della A.16 si ha che le funzioni:

$$\phi_{r,m}(x) = \sqrt{2^r} \phi(2^r x - m), \quad (\text{A.18})$$

con $m \in \mathbb{Z}$, costituiscono una base ortonormale di V_r . Dalle condizioni A.11 e A.16 discende che $\phi \in V_0 \subset V_1$, perciò $\phi(x)$ può essere espresso come combinazione lineare degli elementi della base di V_1 :

$$\phi(x) = \sum_{m \in \mathbb{Z}} h_m \phi_{1,m} = \sqrt{2} \sum_{m \in \mathbb{Z}} h_m \phi(2x - m). \quad (\text{A.19})$$

Questa equazione prende il nome di *equazione di dilatazione* e i pesi della combinazione lineare $h_m = \langle \phi | \phi_{1,m} \rangle$ sono detti *coefficienti di filtro*; si dimostra che le condizioni A.11-A.16 stabiliscono per l'insieme dei coefficienti di filtro le seguenti relazioni:

$$\sum_{m=-\infty}^{+\infty} h_m = \sqrt{2}, \quad (\text{A.20})$$

$$\sum_{m=-\infty}^{+\infty} (-1)^m h_m = 0; \quad (\text{A.21})$$

la condizione di ortonormalità impone inoltre:

$$\sum_{m \in \mathbb{Z}} h_m \overline{h_{2k+m}} = \delta_{m0}. \quad (\text{A.22})$$

Data $f \in L_2(\mathbb{R})$, una sua approssimazione alla *risoluzione* 2^r è definita come la sua proiezione ortogonale nello spazio V_r :

$$f_r(x) = \sum_{m \in \mathbb{Z}} \langle f | \phi_{r,m} \rangle \phi_{r,m}(x). \quad (\text{A.23})$$

La condizione A.12 assicura che $\lim_{r \rightarrow +\infty} f_r(x) = f(x)$ per ogni $f \in L_2(\mathbb{R})$, mentre dalla A.11 si deduce che per valori decrescenti di r si ottengono approssimazioni sempre peggiori, ma più lisce e regolari. I *dettagli* che vengono aggiunti nel passare dallo spazio V_r a V_{r+1} appartengono allo spazio W_r , complemento ortogonale di V_r in V_{r+1} :

$$V_{r+1} = V_r \oplus W_r. \quad (\text{A.24})$$

Iterando la relazione A.24 e sfruttando le condizioni A.12 ed A.13 si riesce a dimostrare che:

$$L_2(\mathbb{R}) = \bigoplus_{r \in \mathbb{Z}} W_r = V_{r_0} \oplus \bigoplus_{r \geq r_0} W_r. \quad (\text{A.25})$$

A questo punto si consideri una funzione $\psi(x) \in W_0$, detta *wavelet madre* tale che $\{\psi(x - m), m \in \mathbb{Z}\}$ sia una base ortonormale in W_0 ; ne consegue che le funzioni:

$$\psi_{r,m}(x) = \sqrt{2^r} \psi(2^r x - m), \quad (\text{A.26})$$

costituiscono una base ortonormale per W_r e quindi. Questa espressione coincide con la A.9 a patto di sostituire l'indicizzazione in r con quella equivalente in $j = -r$

Tenendo conto della relazione A.25, per ogni funzione $f(x) \in L_2(\mathbb{R})$ valgono dunque le seguenti *rappresentazioni wavelet*:

$$f(x) = \sum_{r,m \in \mathbb{Z}} d_{r,m} \psi_{r,m}(x) = \sum_{m \in \mathbb{Z}} c_{r_0,m} \phi_{r_0,m}(x) + \sum_{r=r_0}^{+\infty} \sum_{m \in \mathbb{Z}} d_{r,m} \psi_{r,m}(x). \quad (\text{A.27})$$

con:

$$c_{r,m} = \langle f | \phi_{r,m} \rangle = \int_{\mathbb{R}} f(x) \overline{\phi_{r,m}(x)} dx, \quad (\text{A.28})$$

$$d_{r,m} = \langle f | \psi_{r,m} \rangle = \int_{\mathbb{R}} f(x) \overline{\psi_{r,m}(x)} dx, \quad (\text{A.29})$$

I coefficienti $c_{r,m}$ e $d_{r,m}$ sono detti rispettivamente *coefficienti di scala* e *coefficienti wavelet*.

Deve essere sottolineato che, in effetti, un'analisi a più livelli di risoluzione è sostanzialmente definita dalla scelta della funzione di scala ϕ . Infatti poiché $W_0 \subset V_1$ la wavelet madre ψ può essere espressa come combinazione lineare degli elementi della base di V_1 :

$$\psi(x) = \sum_{m \in \mathbb{Z}} g_m \phi_{1,m}(x) = \sqrt{2} \sum_{m \in \mathbb{Z}} g_m \phi(2x - m). \quad (\text{A.30})$$

I coefficienti g_m possono essere scelti imponendo la condizione:

$$g_m = (-1)^m \overline{h_{1-m}}, \quad (\text{A.31})$$

che risulta essere sufficiente a garantire l'ortonormalità della base wavelet che ne discende.

Storicamente la wavelet di Haar, [26], è stata introdotta e studiata ben prima che venisse formalizzata la teoria delle wavelet. Si tratta di una wavelet ortonormale, analiticamente rappresentata dalla funzione:

$$\psi(x) = \begin{cases} +1, & \text{se } x \in [0, \frac{1}{2}); \\ -1, & \text{se } x \in [\frac{1}{2}, 1); \\ 0, & \text{altrimenti;} \end{cases} \quad (\text{A.32})$$

che si ottiene assumendo come funzione di scala la funzione *boxcar*:

$$\phi(x) = I_{[0,1]}(x) = \begin{cases} 1, & \text{se } x \in [0, 1); \\ 0, & \text{altrimenti.} \end{cases} \quad (\text{A.33})$$

Una proprietà notevole di questa wavelet è quella di risultare nulla al di fuori di un intervallo limitato di x , ovvero di essere definita con un numero finito di coefficienti di filtro non nulli: $h_0 = h_1 = g_0 = -g_1 = \frac{1}{\sqrt{2}}$.

Anche le famiglie di wavelets ortogonali costruite da Daubechies, [14], [15], [16], godono della proprietà di avere supporto compatto e un numero finito di coefficienti di filtro; si distinguono per altre condizioni che vengono imposte, come un'asimmetria minima o un certo numero di momenti nulli. È interessante osservare che le wavelets di Daubechies, pur essendo le più utilizzate nelle applicazioni, non hanno una formulazione analitica e sono definite soltanto in termini di coefficienti di filtro.

Naturalmente esistono anche wavelet con supporto esteso sull'intero asse reale, anche se in ogni caso l'andamento deve tendere a zero per $|x| \rightarrow +\infty$. In questi casi si hanno infiniti coefficienti di filtro non nulli, ma, stabilita una certa tolleranza, soltanto un numero finito di essi può essere considerato non trascurabile.

Inoltre è doveroso sottolineare come non sia corretto restringere l'attenzione soltanto a sistemi wavelet ortogonali. Nella trasformata wavelet continua non ha senso imporre condizioni di ortogonalità alla wavelet madre ψ , dal momento che in ogni caso la base che si utilizza risulta sovracompleta. Si preferisce pertanto sostituire la condizione di ortogonalità con qualche altra condizione, da imporre nella scelta della wavelet madre affinché la trasformata ottenuta abbia qualche proprietà desiderabile nel contesto applicativo in cui si opera. Ad esempio in [21] vengono discussi i vantaggi dell'impiego della wavelet di Morlet, che oltre ad essere non ortogonale è anche a valori complessi. Un altro popolare esempio di wavelet non ortogonale è dato dalla *mexican hat*, proporzionale alla derivata seconda della funzione gaussiana $e^{-x^2/2}$.

Anche nel contesto della trasformata wavelet discreta l'ortogonalità non è un obbligo: vengono comunemente utilizzate delle basi wavelet non ortogonali che sono dette *biortogonali*. Si tratta di una coppia di basi wavelet coniugate utilizzate l'una in sede di analisi, cioè nella trasformata diretta, e l'altra in sede di sintesi, cioè nella trasformata di ricostruzione. Le due basi sono vincolate

l'una all'altra, ma consentono molta più flessibilità rispetto ad un sistema ortogonale: ad esempio le wavelet di Vetterli ed Herley, [43] sono simmetriche pur avendo supporto compatto, mentre tra i sistemi ortogonali a supporto compatto si riesce ad imporre soltanto un vincolo di asimmetria minima, [15].

Bibliografia

- [1] **Adaptive signal processing**
Alexander S.T., (1986), Springer-Verlag, New York

- [2] **Methods for curve estimation**
Antoniadis A., Gregoire G., McKeague I.W., (1994), *J. Amer. Statist. Assoc.*, **89**, 428

- [3] **Wavelets and statistics**
Antoniadis A. & Oppenheim G. (ed.), (1995), *Lecture notes in statistics*, **103**, Springer-Verlag, New York

- [4] **Two filtering methods for the detection of gravitational wave bursts**
Arnaud N., Cavalier F., Davier M., Hello P., (1998), *VIRGO Note*, **110**

- [5] **One peak filtering method for the detection of gravitational wave bursts**
Arnaud N., Cavalier F., Davier M., Hello P., (1998), *VIRGO Note*, **118**

- [6] **Study of a three dimensional set of filters**
Arnaud N., Cavalier F., Davier M., Hello P., (1998), *VIRGO Note*, **125**

- [7] **Time series - Data analysis and theory**
Brillinger D.R., (1981), Holden Day Inc., San Francisco

- [8] **Wavelet detection of transients in noisy time series**
Carmona R., (1992), *Technical Report*, Department of Mathematics
University of California at Irvine

- [9] **Triggers for the detection of impulsive sources of gravitational waves**
Cavalier F., Davier M., Hello P., Pradier T., (1999), *VIRGO Note*, **128**
- [10] **An introduction to wavelets**
Chui C.K., (1992), *Wavelet analysis and its applications*, **1**, Academic Press, San Diego
- [11] **Entropy-based algorithms for best basis selection**
Coifman R.R. & Wickerhauser M.V., (1992), *IEEE Trans. Inf. Theory*, **38**, 713
- [12] **Detection of turbulent coherent motions in a forest canopy. Part I: wavelet analysis**
Collineau S. & Brunet Y., (1993), *Boundary-Layer Meteorology*, **65**, 357
- [13] **Identificazione dello spettro di rumore di VIRGO e metodi di rivelazione dei segnali gravitazionali**
Cuoco E., (1997), *Tesi di Dottorato in Fisica*, Dipartimento di Fisica Università degli Studi di Pisa
- [14] **Orthonormal bases of compactly supported wavelets**
Daubechies I., (1988), *Comm. Pure Appl. Math.*, **41**, 909
- [15] **Ten lectures on wavelets**
Daubechies I., (1992), *CBMS-NFS Series in Applied Mathematics*, **61**
- [16] **Orthonormal bases of compactly supported wavelets II. Variations on the theme**
Daubechies I., (1993), *SIAM J. Math. Anal.*, **24**, 499
- [17] **Ideal spatial adaption via wavelet shrinkage**
Donoho D.L. & Johnstone I.M., (1994), *Biometrika*, **81**, 425
- [18] **Adapting to unknown smoothness by wavelet shrinkage**
Donoho D.L. & Johnstone I.M., (1995), *J. Amer. Statist. Assoc.*, **90**, 1200
- [19] **Wavelet shrinkage: asymptotia?**
Donoho J.L., Johnstone I.M., Kerkyacharian G., Picard D., (1995), *J. R. Statis. Soc. B*, **57**, 301

- [20] **Estimating covariances of locally stationary processes: rates of convergence of best basis methods**
Donoho D.L., Mallat S., von Sachs R., (1998), *Annals of Statistics*, subm.
- [21] **Wavelet transforms and their applications to turbulence**
Farge M., (1992), *Annual Rev. Fluid Mech.*, **24**, 395
- [22] **Time-frequency / time-scale analysis**
Flandrin P., (1999), Academic Press, San Diego
- [23] **Transient signal detection using prior information in the likelihood ratio test**
Frisch M. & Messer H., (1993), *IEEE Trans. Signal Proc.*, **41**, 2177
- [24] **Wavelet estimation of spectral densities in time series analysis**
Gao H., (1993), *PhD Thesis*, University of California, Berkeley
- [25] **Choice of thresholds for wavelet shrinkage estimate of the spectrum**
Gao H., (1997), *Journal of Time Series Analysis*, **18**, 231
- [26] **Zur Theorie der orthogonalen Funktionensysteme**
Haar A., (1910), *Mathematische Annalen*, **69**, 331
- [27] **Time series analysis**
Hamilton J.D., (1994), Princeton University Press, Princeton
- [28] **Pulse events - Detection of impulsive sources**
Hello P. (coord.), (1999), in *VIRGO Data Analysis Documents*
- [29] **The effect of rotation on the gravitational radiation and dynamical stability of stiff stellar cores**
Houser J., (1999), *Mon. Not. Roy. Astron. Soc.*, **299**, 1069
- [30] **Wavelets and binary coalescences detection**
Innocent J.M. & Torresani B., (1997), *Applied and Computational Harmonic Analysis*, **4**, 113
- [31] **An overview of wavelet based multiresolution analyses**
Jawerth B. & Sweldens W., (1994), *SIAM Review*, **36**, 377

- [32] **Wavelet threshold estimators for data with correlated noise**
Johnstone I.M. & Silverman B.W., (1997), *J. R. Statist. Soc. B*, **59**, 319
- [33] **Density estimation by kernel and wavelets methods**
Kerkycharian G. & Picard D., (1993), *Statistics & Probability Letters*, **12**, 327
- [34] **Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$**
Mallat S., (1989), *Trans. Amer. Math. Soc.*, **315**, 69
- [35] **Simultaneous statistical inference**
Miller R.G., (1981), Springer-Verlag, New York
- [36] **Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum**
Nason G.P., von Sachs R., Kroisandt G., (2000), *J. R. Statist. Soc. B*, **62**, 271
- [37] **Introduction to signal processing**
Orfanidis S.J., (1996), Prentice Hall, Englewood Cliffs (NJ)
- [38] **Spectral analysis for physical applications. Multitaper and conventional univariate techniques**
Percival D.B. & Walden A.T., (1993), Cambridge University Press, Cambridge
- [39] **Wavelet methods for time series analysis**
Percival D.B. & Walden A.T., (2000), Cambridge University Press, Cambridge
- [40] **Spectral analysis and time series**
Priestley M.B., (1987), Academic Press, New York
- [41] **Time-frequency and time-scale analysis**
Torresani B., (1999), in *Signal processing for multimedia*, Byrnes J. (ed.)
- [42] **Detection, estimation and modulation theory**
Van Trees H.L., (1971), Wiley, New York

- [43] **Wavelets and filter banks: theory and design**
Vetterli M. & Herley C., (1992), *IEEE Trans. Signal Processing*, **40**, 2207
- [44] **Statistical modeling by wavelets**
Vidakovic B., (1999), John Wiley and sons, New York
- [45] **The VIRGO interferometer**
VIRGO collaboration, (1997), *Classical and quantum gravity*, **14**, 1461
- [46] **Automatic smoothing of the log-periodogram**
Wahba G., (1980), *J. Amer. Statist. Assoc.*, **75**, 122
- [47] **Probabilities for the size of largest clusters and smallest intervals**
Wallenstein S. & Naus J., (1974), *J. Amer. Statist. Assoc.*, **68**, 690
- [48] **An approximation for the distribution of the scan statistic**
Wallenstein S. & Neff N., (1987), *Statistics in Medicine*, **6**, 197
- [49] **Bayesian forecasting and Dynamic Models**
West M. & Harrison J., (1989), Springer-Verlag, New York
- [50] **An improved Bonferroni inequality and applications**
Worsley K.J., (1982), *Biometrika*, **69**, 297
- [51] **Wavelet theory and applications**
Young R., (1993), Kluwer Academic Publisher
- [52] **Dynamics and gravitational wave signature of axisymmetric rotational core collapse**
Zwenger T. & Müller E., (1997), *Astron. Astrop.*, **320**, 209