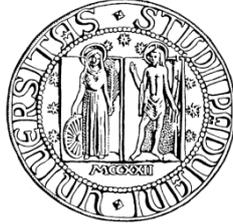


UNIVERSITÀ DEGLI STUDI DI PADOVA



Facoltà di Ingegneria

Dei-Dipartimento di Ingegneria dell'informazione

Corso di Laurea Specialistica in Bioingegneria

Modelli della regolazione genica nel lievito

Relatrice: **Prof.ssa Gianna Maria Toffolo**

Correlatrice: **Federica Eduati**

Laureanda: **Anna Pompolani**

ANNO ACCADEMICO 2009-2010

Sommario

In risposta ad uno stimolo le cellule modificano l'espressione genica per adattarsi alle nuove condizioni. Dai dati sperimentali si è notato (Gasch et al. 2000) che un elevato numero di geni del *S. Cerevisiae* risponde con un andamento stereotipato che è stato definito Environmental Stress Response (ESR) caratterizzato da un rapido picco iniziale seguito dal ritorno al livello basale. Utilizzando i dati di espressione genica è stato dimostrato (Altafini et al. 2009) che è possibile descrivere in modo qualitativo il pattern ESR con un modello "feedback-integrativo" in cui un aumento nell'abbondanza di una certa proteina sfavorisce la trascrizione del corrispondente gene. Durante il lavoro di tesi è stato definito un modello diverso pensando che la regolazione del gene avvenga ad opera dell'azione combinata di altri due geni secondo il meccanismo descritto dal "feed-forward loop" che è uno dei pattern di interconnessione significativamente frequenti all'interno delle reti reali.

Sono stati implementati i due modelli in ambiente Matlab e ne è stata fatta l'identificazione parametrica usando i dati di espressione di un set di geni del lievito caratterizzati dalla tipica risposta allo stress. Ci si è posti come obiettivo quello di capire quale dei modelli descriva meglio il profilo considerato.

Indice

Introduzione	7
1.1. Environmental Stress Response (ESR)	7
1.2. Modellizzazione dell'ESR	9
I dati.....	11
2.1. Espressione genica.....	11
2.1.1 Tecniche di misura.....	13
2.1.2 Saccharomyces cerevisiae in risposta allo stress	17
2.2. Reti di regolazione.	21
2.2.1 Rete trascrizionale di Saccharomyces cerevisiae e FFL.....	25
2.3. Associazione dei dati.....	27
I modelli	35
3.1. Reverse engineering	35
3.2 Costruzione dei modelli	39
3.2.1 Modello "Feedback integrativo"	40
3.2.2 Modello "Feed-forward loop"	42
3.3. Implementazione dei modelli	44

Identificazione dei modelli.....	53
4.1. Modello “feed-forward”	54
4.1.1 Stima di k_{31} e k_{32}	54
4.1.2. Stima di k_{31} , k_{32} e k_{03}	68
4.2. Modello “feedback-integrativo”	77
4.2.1. Stima di a_i e b_i	77
4.2.2. Stima di a_i , b_i e δ_i	83
Risultati	91
5.1. Risultati.....	91
Conclusioni.....	101
6.1. Conclusioni	101
Bibliografia	103

Capitolo 1

Introduzione

1.1. Environmental Stress Response (ESR)

Gli organismi cellulari necessitano di specifiche condizioni interne per consentire una crescita e un funzionamento ottimali. I diversi organismi hanno sviluppato molteplici strategie per mantenere queste condizioni interne a fronte di cambiamenti, spesso violenti e repentini, dell'ambiente esterno. Mentre gli organismi pluricellulari possono utilizzare tessuti ed organi specializzati per garantire delle condizioni interne relativamente stabili e omogenee, gli organismi unicellulari, come il lievito *Saccharomyces Cerevisiae*, hanno sviluppato dei meccanismi autonomi per permettere l'adattamento ai drastici cambiamenti ambientali.

I lieviti regolarmente resistono a fluttuazioni nel tipo e nella quantità di nutrienti disponibili, a variazioni di temperatura, osmolarità e acidità dell'ambiente in cui vivono e alla presenza variabile di agenti nocivi come radiazioni e prodotti chimici tossici. Ecco che l'espressione genica che consente alla cellula di rispondere al meglio alle diverse situazioni sarà differente a seconda delle condizioni ambientali in cui si trova l'organismo. Quindi, quando le condizioni ambientali cambiano rapidamente, la cellula deve repentinamente aggiustare la sua espressione genica per adattarsi alle nuove condizioni.

Nel lavoro di Gasch [1] sono stati esplorati i pattern di espressione genica nel lievito *Saccharomyces cerevisiae* ottenuti in risposta a differenti stimoli. Sono stati usati i microarray di DNA per analizzare i cambiamenti nei livelli di espressione per circa 6200 geni del lievito in risposta a vari tipi di stress ambientali tra cui shock termici, trattamento con perossido di idrogeno, esposizione a menadione sodio bisolfito e altri. Una caratteristica sorprendente rivelata da questa analisi è che un elevato numero di geni, circa 900 (più del 14% dei geni attualmente identificati nel genoma del lievito), risponde con un andamento stereotipato a ciascuna delle condizioni di stress che sono state testate. Questo tipo di risposta, definita "Environmental stress

response" (ESR), è costituita da un rapido picco di espressione (nell'ordine della decina di minuti) seguito da un ritorno al livello pre-stimolo. Tra i geni caratterizzati da ESR si individuano 2 grandi cluster che mostrano profili temporali reciproci ma pressoché identici nell'andamento: uno relativo ai geni repressi e uno relativo ai geni indotti. I geni repressi (circa 600) possono essere a loro volta distinti in 2 gruppi in base al profilo di espressione: i geni del primo gruppo sono collegati al processo di crescita cellulare, a vari aspetti del metabolismo dell'RNA (processamento e splicing RNA, iniziazione ed elongazione della traduzione, sintesi e processamento del tRNA), alla biosintesi dei nucleotidi, alla secrezione e ad altri processi metabolici. I geni del secondo gruppo sono caratterizzati da un lieve ritardo nel declino del livello di espressione dopo lo stimolo e si tratta per lo più di geni che codificano per proteine ribosomiali. Il cluster dei geni indotti raggruppa circa 300 geni legati a molteplici processi tra i quali il metabolismo dei carboidrati, le reazioni di ossido-riduzione e di difesa dalle specie reattive dell'ossigeno, la modificazione delle pareti cellulari, il ripiegamento e la degradazione delle proteine, il trasporto di metaboliti.

Considerando le funzioni a cui sono legati i geni che presentano andamento ESR possiamo supporre quali possono essere le caratteristiche che la cellula tenta di preservare in risposta allo stress. La necessità di conservazione dell'energia è molto probabilmente un'esigenza che caratterizza tutte le risposte allo stress e il tipico andamento ESR presumibilmente aiuta a raggiungere tale scopo attraverso la repentina disattivazione della trascrizione di centinaia di geni coinvolti nei processi di sintesi proteica e di crescita cellulare. I processi a cui sono collegati i geni indotti fanno supporre che il ruolo della risposta allo stress, in questo caso, sia quello di provvedere alla fortificazione della cellula e di garantire la sua stessa sopravvivenza in un ambiente inospitale [1]. La rapida e significativa escursione transitoria è dunque d'aiuto nell'attivazione immediata dei meccanismi di reazione cellulare (come ad esempio la redistribuzione delle risorse energetiche), mentre l'adattamento, cioè il riassorbimento dell'escursione, diventa uno strumento per ripristinare un punto di lavoro il più possibile vicino all'ottimo a dispetto di cambiamenti ambientali più o meno persistenti [3]. Si intuisce dunque l'importanza che questa risposta assume all'interno del fine e complesso sistema globale di regolazioni che permette alla cellula di svolgere molteplici attività garantendo la sua sopravvivenza, la sua crescita e delle prestazioni specifiche e ottimali.

1.2. Modellizzazione dell'ESR

Data l'importanza e la peculiarità dell'adattamento dell'organismo a nuove condizioni ambientali, diversi studi si sono concentrati sull'analisi e la modellizzazione del pattern che caratterizza l'ERS nel lievito. Tra questi studi citiamo il lavoro di Koller [2] in cui si è trovato che, nel metabolismo del lievito, il tipico andamento è il risultato di una regolazione trascrizionale finemente sincronizzata che si innesca con lo stimolo. Il suddetto lavoro si è focalizzato sullo studio della dinamica delle reti e ha portato all'identificazione dei cosiddetti "timing activity motifs" ovvero particolari circuiti in grado di descrivere gli specifici meccanismi temporali che regolano l'ordine di attivazione dei componenti della rete. Il lavoro di Koller analizza dunque il profilo ESR cercando di catturare la fine regolazione temporale che causa una precisa sincronizzazione tra i componenti stessi.

Un modello matematico in grado di spiegare l'andamento che caratterizza l'ERS è stato presentato da Altafini [3] e parte dall'osservazione secondo la quale, affinché l'espressione genica ritorni al suo livello basale indipendentemente dall'ampiezza dello stimolo, il sistema deve avere un robusto meccanismo di regolazione oltre a possedere una memoria del valore di concentrazione nominale pre-stimolo per ogni gene. Altafini propone un modello capace di spiegare questa rapida risposta con un meccanismo a retroazione integrale finalizzato a far adattare il sistema alla nuova condizione. In questo schema l'integrale della distanza dal valore nominale (quindi l'errore) di una variabile costituisce il ramo in retroazione con segno negativo. L'adattamento viene raggiunto appena la variabile ritorna al valore nominale (cioè l'errore è zero) a fronte di uno stimolo persistente che, in assenza di feedback, altererebbe il valore dello stato stazionario. La variabile che viene integrata è l'abbondanza relativa di mRNA e il suo integrale rappresenta l'abbondanza relativa del corrispondente prodotto genico. Si assume che questa quantità agisca omeostaticamente sulla velocità di trascrizione dell'mRNA, riportando l'espressione genica al valore nominale di concentrazione.

L'autoregolazione negativa della trascrizione è un meccanismo che permette di ridurre le fluttuazioni attorno allo stato stazionario e di diminuire il tempo di salita della risposta. Il modello proposto da Altafini permette quindi di descrivere le principali caratteristiche dell'ERS basandosi su un meccanismo di autoregolazione in cui è la proteina stessa, prodotta a partire dall'mRNA, a fungere da regolatore dell'espressione del gene che la codifica.

Dalla letteratura è noto però, che differenti geni e prodotti genici interagiscono tra loro andando a formare una complessa rete di regolazione. Tra geni, mRNA, proteine e metaboliti si instaurano molteplici e differenti interazioni e meccanismi di controllo che permettono di regolare l'espressione dei geni in base alle necessità della specifica cellula in ogni istante di

tempo. Si capisce quindi come sia estremamente importante tener conto del sistema nel suo complesso per poter inferire sulla topologia e sulla dinamica della rete di regolazione trascrizionale e delle sue singole componenti.

Il presente lavoro di tesi parte, quindi, dal modello proposto da Altafini e punta allo sviluppo di un modello alternativo che tenga in considerazione anche la rete di regolazione trascrizionale. Per raggiungere tale obiettivo si è pensato di focalizzare l'attenzione su particolari pattern di interconnessione significativamente frequenti, chiamati "motivi di rete" o "network motifs".

È noto infatti che le reti possono essere scomposte in circuiti elementari alcuni dei quali sono presenti in una rete un numero di volte statisticamente rilevante il che suggerisce che abbiano un significato biologico ben preciso. Gli stessi motivi di regolazione sono stati trovati in organismi diversi, batteri, piante e animali e ciò suggerisce che essi siano stati selezionati grazie alle loro proprietà durante l'evoluzione, in quanto vantaggiosi per l'organismo. Diventa allora molto interessante caratterizzare in termini teorici le proprietà di questi moduli elementari di regolazione utilizzando dei semplici modelli matematici.

Modelli teorici ed esperimenti su sistemi reali consentono di indagare sulle specifiche funzioni che questi motivi possono realizzare.

In questo lavoro di tesi viene proposto un modello per interpretare il tipico profilo ESR, pensando che possa essere proprio uno di questi motivi della rete a provocarne l'attivazione. Il motivo da noi considerato è il "feed-forward loop", nel quale un fattore di trascrizione regola un secondo fattore di trascrizione e la loro azione combinata regola l'espressione di un gene target. Si può infatti pensare che lo stimolo esterno provochi l'innescamento di alcuni segnali che vanno ad attivare o ad inibire i fattori di trascrizione i quali regolano la trascrizione del gene target.

Poiché questo motivo agisce da acceleratore della risposta trascrizionale (attivazione o repressione) del gene target, può essere utilizzato per spiegare il meccanismo tramite il quale la cellula fa fronte in modo repentino ai cambiamenti delle condizioni esterne.

Capitolo 2

I dati

2.1. Espressione genica.

In ogni cellula somatica di ogni organismo eucariote è contenuto tutto il genoma, ovvero tutta l'informazione necessaria all'organismo stesso per poter sopravvivere, codificata all'interno del DNA o codice genetico. L'informazione genica specifica la natura e le proprietà delle proteine e delle molecole funzionali che l'organismo è in grado di produrre.

Il passaggio dell'informazione codificata in un gene, prima all'RNA messaggero e poi alla proteina (o al solo RNA se il gene codifica per un RNA funzionale) è detto "espressione genica". L'espressione genica diversificata rende possibile la differenziazione cellulare, lo sviluppo, l'attività cellulare, la risposta e l'adattamento agli stimoli esterni. Infatti le cellule non contengono solo le istruzioni per la codifica delle proteine, ma anche l'informazione relativa alle condizioni in cui le proteine devono essere sintetizzate.

Questa informazione si esplica attraverso meccanismi di regolazione e controllo molto complessi, i cui due passaggi principali sono la trascrizione, durante la quale il DNA è trascritto in RNA, e la traduzione, durante la quale l'RNA è tradotto in proteina.

La regolazione della trascrizione si può spiegare attraverso un modulo che presenta quattro elementi costitutivi (Figura 2.1): il promotore, i geni espressi dalla sua attivazione, le proteine regolatrici e le molecole che modulano la loro attività (induttori). L'interazione a livello molecolare tra questi elementi determina il livello di trascrizione dei geni.

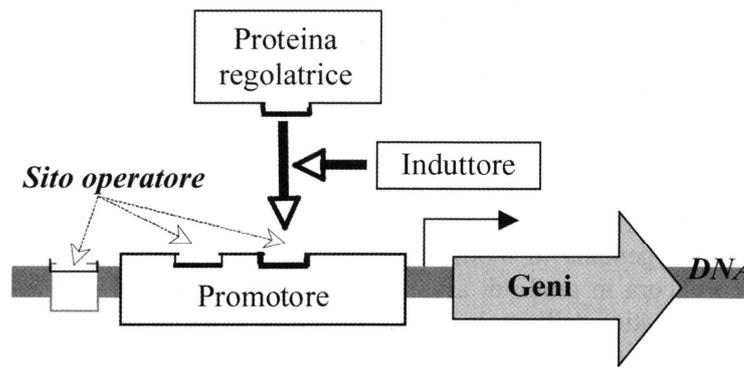


Figura 2.1 Architettura di un modulo di regolazione trascrizionale.

Il controllo dell'espressione genica è operato dalle proteine regolatrici, chiamate fattori di trascrizione, che sono in grado di interagire con una breve porzione di DNA non codificante (promotore) che fiancheggia la sequenza da trascrivere. Un fattore di trascrizione interagisce con il DNA legandosi a siti specifici detti siti operatore, che si trovano nella regione del promotore, e causa un aumento (attivatore) o una diminuzione (repressore) della probabilità che i geni regolati siano trascritti in mRNA. Di fatto, un fattore di trascrizione si comporta da attivatore quando aumenta la probabilità che l'RNA polimerasi leghi il promotore per iniziare la trascrizione e di conseguenza un attivatore fa aumentare il tasso di trascrizione. Il contrario accade nel caso del repressore. Ovviamente, la probabilità che un repressore o un attivatore leghino il DNA dipende dalla loro affinità e dalla loro concentrazione. Un repressore, se presente in concentrazione sufficiente, può silenziare completamente la trascrizione. Viceversa, quando il promotore è attivo, esiste un tasso massimo di trascrizione, quindi una saturazione del livello di espressione, determinato dalla concentrazione limitata di RNA polimerasi.

Negli organismi superiori, dopo la trascrizione, alcune regioni non codificanti dell'RNA trascritto, chiamate introni, vengono rimosse dalla sequenza nucleotidica (processo noto come "splicing"); la sequenza rimanente, detta RNA messaggero (mRNA) viene trasportata fuori dal nucleo e tradotta in proteina grazie all'interazione con i ribosomi ed altre molecole.

Sebbene il passo più efficiente in termini di regolazione avvenga a livello della trascrizione, i processi di splicing, di eventuale degradazione della molecola di RNA, di trasporto fuori dal nucleo e di riconoscimento da parte dei ribosomi, vengono regolati da diversi tipi di proteine e RNA.

Infine esiste un sistema di controllo post-traduzionale che regola le interazioni proteina-proteina e le modifiche post-traduzionali (come la fosforilazione), in altre parole tutta una serie di trasformazioni e interazioni che possono modificare l'attività e la funzione della proteina e quindi, ad esempio, la sua capacità di regolare la trascrizione di altri geni.

È chiaro allora che è l'espressione genica, e non il solo DNA, a dire quale sia lo stato di una particolare cellula in un dato momento.

Un gene si dice quindi espresso se nella cellula sono presenti l'RNA o le proteine che vengono codificate da esso o entrambi, quindi, per poter valutare l'espressione genica di una cellula in un preciso istante, dobbiamo valutare quali siano le concentrazioni di RNA nella cellula stessa oppure la concentrazione delle proteine o entrambe. Generalmente però le tecnologie per l'analisi delle proteine sono fortemente limitate rispetto all'impiego in quest'ambito, infatti raramente sono in grado di dirci quali siano le proteine coinvolte, e quindi spesso si ricorre al solo livello di espressione dell'RNA.

Ma come si misura la concentrazione di RNA nella cellula? Di seguito vengono presentati due diversi tipi di tecnologia impiegati per la misurazione di tale concentrazione.

2.1.1 Tecniche di misura

Northern Blotting

Si tratta della prima tecnica utilizzata (1977) per il riconoscimento dell'espressione di specifiche sequenze di DNA note. Per sequenza di DNA nota si intende che sono noti la sequenza nucleotidica oppure la posizione nel genoma, o entrambi.

L'RNA, in questa tecnica, viene prelevato dai campioni, pulito, sottoposto a elettroforesi su gel e quindi fissato su un supporto costituito da una membrana di nylon. L'RNA prende il nome di "target".

La sequenza di DNA o sonda (probe) corrispondente al gene che voglio monitorare viene prelevata dal DNA (o sintetizzata in laboratorio), replicata (ad esempio tramite PCR), marcata e messa a contatto con l'RNA immobilizzato sulla membrana di nylon.

Per via della proprietà di appaiamento delle sequenze di basi che compongono i due acidi nucleici, se il gene è espresso nel campione allora il DNA si fisserà all'RNA complementare e in tale caso si parla di ibridazione. Il marcatore permette di identificare quali frammenti di DNA si sono legati.

Si tratta di una tecnologia che non dà informazioni se non qualitative di quali siano le sequenze di DNA espresse.

Microarray

A differenza del Northern Blotting, i microarray permettono di monitorare contemporaneamente decine di migliaia di geni; sono quindi tecnologie high throughput che potenzialmente permettono di monitorare l'espressione di tutti i geni di un genoma. Questo è possibile perché quella dei microarray è una tecnologia ad alta densità, che permette di ancorare su un supporto di piccole dimensioni migliaia di sonde diverse.

Il principio di funzionamento dei microarray si basa anch'esso sull'ibridazione DNA-RNA resa possibile dalla complementarità delle basi, tuttavia, a differenza del Northern Blotting, nei microarray il DNA (probe) viene ancorato su un supporto solido (in vetro o silicio) e l'RNA target prelevato da una cellula è libero di legarsi al DNA ad esso complementare. Analizziamo ora le due tipologie fondamentali e di maggior utilizzo di microarray.

Microarray spot a due canali

Le sonde (probe) sono sequenze di DNA gene-specifiche di un organismo (in genere lunghe 500-5000 bp) che vengono opportunamente scelte attraverso la consultazione di banche dati genomiche e vengono replicate per aumentarne il numero.

Un robot preleva le sonde dalle piastre dove sono conservate e, con precisione micrometrica, le deposita su un supporto (tipicamente un vetrino da microscopia) attraverso un sistema di pennini. Si ottiene così un micro-schieramento ordinato di sonde, da cui il nome microarray, che viene trattato chimicamente in modo da immobilizzare le sonde e da denaturare il DNA per renderlo a singolo filamento.

L'area coperta da ogni punto stampato prende il nome di "spot" e contiene catene di DNA del medesimo tipo di cui vengono memorizzate la posizione e la particolare sequenza. Con questa tecnica si possono avere circa 10^4 molecole di probe per ogni spot.

Dalle cellule di interesse si estrae l'RNA target e da questo si sintetizza il DNA complementare (cDNA) mediante l'uso dell'enzima trascrittasi inversa. Nel corso della reazione di retro-trascrizione il cDNA viene marcato con molecole fluorescenti e messo a contatto per qualche ora con il microarray. Il cDNA ibrida ovvero si accoppia in maniera complementare e

specifica alle catene di DNA depositate sul vetrino (si dice che il microarray viene interrogato). Dopo 8-10 ore il vetrino è sottoposto ad opportuni lavaggi per eliminare le molecole che non hanno ibridato.

Se ora si eccita il fluorescente con una luce laser, esso emetterà onde elettromagnetiche a lunghezza d'onda caratteristica. L'intensità dell'emissione per ciascuno spot, rilevata da uno scanner ad alta sensibilità, è una misura di quante molecole hanno ibridato la probe corrispondente, ovvero di quanto il gene associato alle probe è espresso nella cellula.

Vi sono problematiche nell'utilizzo di un solo canale di RNA per ibridare un microarray di questo tipo. Ad esempio si possono verificare facilmente fenomeni di saturazione dato che per ogni spot, quindi per ogni gene monitorato, il numero di probe, cioè di molecole ancorate al vetrino, è di circa 10^4 . La saturazione impedisce di ottenere un'informazione precisa dell'espressione.

Inoltre, date le differenti specificità delle probe per i geni che devono sondare, non si possono fare confronti tra i segnali di probe diverse che possono avere ibridazioni diverse a parità di espressione.

Queste problematiche vengono risolte mediante l'utilizzo di due campioni di RNA (per esempio paziente e controllo) sottoposti a ibridazione competitiva sullo stesso microarray. In questo caso la tecnologia prende il nome di microarray spot a due canali.

Uno dei due campioni è noto ed è usato come test per ottenere delle misure relative di espressione del campione da esaminare, detto "reference".

Per poter misurare separatamente i due segnali, "test" e "reference", i due campioni sono retrotrascritti in cDNA e marcati con due molecole fluorescenti differenti. Solitamente i due fluorofori usati sono due cianine: per il reference la cianina 3 (Cy3), che emette a una lunghezza d'onda nella gamma del verde, e per il test la cianina 5 (Cy5) che emette nella gamma del rosso.

Dopodiché i campioni possono essere mescolati e posti ad ibridare insieme su un unico vetrino.

Se nei due campioni marcati con Cy3 e Cy5 sono presenti trascritti di uno stesso gene, essi competeranno per ibridare alle stesse probe.

Uno scanner ad alta sensibilità è utilizzato per rilevare la lunghezza d'onda e l'intensità del segnale. Un software poi sottrae il background all'intensità ottenuta dall'intero vetrino e calcola il rapporto tra i due canali. I valori del rapporto sono poi convertiti in immagine dove a ciascuno spot corrisponde un'intensità di una scala di pseudo-colori con gradazioni del rosso, che indica l'aumentato livello d'espressione nel test rispetto al reference. Viceversa, valori del rapporto negativi, corrispondenti a livelli d'espressione più alti nel test relativamente al reference, sono indicati da varie gradazioni del verde. Il colore giallo è usato per valori del rapporto vicini

all'unità e il nero indica la mancata ibridazione. Ne risulta così un'immagine combinata che riproduce il reticolo del DNA-microarray e consente una visione immediata degli spot con intensità di segnale più elevata in un canale che nell'altro, corrispondente in definitiva al diverso livello di contenuto di RNA messaggero nei campioni ibridizzati.

La strategia a due canali permette di risolvere il problema della saturazione e della specificità delle probe.

Un altro problema è presentato dai fluorofori Cy3 e Cy5 che, nonostante siano molto simili fra loro, presentano una diversa efficienza di fluorescenza che deve necessariamente essere standardizzata dai software di rilevazione.

Inoltre poiché Cy3 è più piccolo di Cy5, esiste un diverso livello di incorporazione dei due fluorofori, in quanto la polimerasi presenta più difficoltà a inserire il nucleotide marcato con Cy5 a causa dell'ingombro sterico.

Un problema che non è risolvibile se non in parte grazie ad un'accurata analisi e scelta delle sequenze delle probe è la "cross-ibridazione", ovvero quel fenomeno che vede ibridare sulle probe catene di target che avrebbero dovuto legarsi alle probe di altri spot.

Microarray di oligonucleotidi

Invece di cDNA è possibile usare catene di nucleotidi (oligonucleotidi) sintetizzate in laboratorio, lunghe circa 70 basi e già a singola catena (si parla di microarray di oligonucleotidi). Nel caso degli Affimetrix Chip® gli oligonucleotidi vengono sintetizzati direttamente su un supporto in silicio (chip) usando tecniche fotolitografiche del tutto simili a quelle utilizzate per costruire i chip per i computer.

Questo permette di avere un numero altissimo di catene per probe, circa 10^9 catene identiche, e il monitoraggio di alcune decine di migliaia di geni contemporaneamente.

Per ogni gene da monitorare vengono utilizzate fino a una ventina di catene diverse, inserite in diverse probe dette "perfect match" (PM), che assieme costituiscono un probe-set per il gene. Per valutare i fenomeni di cross-ibridazione vengono utilizzate delle probe specifiche per ogni PM, dette "mismatch" (MM), che differiscono dagli oligonucleotidi dei PM solo per un nucleotide centrale che quindi non appaia in maniera specifica le sequenze del gene monitorato dal PM.

La soluzione per la misurazione delle concentrazioni in questo caso è diversa, infatti alle catene di RNA che viene retroscritto e poi frammentato (per potersi legare alle brevi sequenze oligonucleotidiche) viene unita una molecola di biotina e non di fluorescente in quanto questa interferisce meno con il processo di ibridazione. Le catene vengono messe a contatto con un

fluorescente che si lega alla biotina solo in un secondo momento per effettuare la lettura dei risultati i quali vanno elaborati con algoritmi specifici per poter essere utilizzati.

2.1.2 *Saccharomyces cerevisiae* in risposta allo stress

Il *Saccharomyces cerevisiae* (lievito della birra), organismo unicellulare appartenente al regno dei funghi, è uno dei microorganismi eucarioti più intensamente studiati in biologia cellulare e molecolare. Nel 1996 è stato il primo eucariote il cui genoma sia stato interamente sequenziato. La banca dati del genoma di *S. cerevisiae* è molto curata e rappresenta un importante strumento di conoscenza basilare delle funzioni e dell'organizzazione della cellula eucariote in genetica e in fisiologia. Il genoma è composto da circa 13.000.000 coppie di basi e 6275 geni, sebbene soltanto 5800 di questi sono ritenuti essere i veri geni funzionali. Si stima che *S. cerevisiae* condivida circa il 23% del suo genoma con l'uomo.

Come abbiamo visto nel capitolo introduttivo, il gruppo di lavoro di Gasch [1] ha esplorato i pattern di espressione genica nel *S. cerevisiae* ottenuti in risposta a differenti stimoli utilizzando gli spot array a due canali descritti nel paragrafo precedente (database disponibile al sito http://genome-www.stanford.edu/yeast_stress).

I profili sono costituiti da un numero variabile di valori di espressione acquisiti in tempi differenti. Ogni valore deriva da un esperimento del tipo stimolo-risposta a cui è stato sottoposto il campione biologico e che ha fornito il livello di espressione di quel gene in quel preciso istante temporale. Più esperimenti dello stesso tipo effettuati in istanti temporali diversi (ovvero usando diversi microarray) hanno dunque fornito un set di valori di espressione che costituisce una serie temporale.

Negli esperimenti l'abbondanza di mRNA (test) viene ibridata contro una concentrazione basale pre-stimolo di mRNA (reference). I 2 campioni vengono marcati rispettivamente con i fluorescenti cianina 5 (Cy5) e cianina 3 (Cy3) e come segnale si considera il logaritmo in base due del rapporto tra le intensità rilevate per i due diversi fluorescenti (si parla di "log-ratio"):

$$LR = \log_2 \left(\frac{S_{Cy5}}{S_{Cy3}} \right) = \log_2(S_{Cy5}) - \log_2(S_{Cy3}) = C - C_{basale}$$

I livelli di espressione misurati dai microarray sono stati organizzati con un clustering gerarchico i cui risultati sono riportati nella Figura 2.2. L'algoritmo di clustering dispone i geni in base alla similarità dei loro profili di espressione ottenuti nei diversi esperimenti, in modo che pattern di

espressione simili siano raggruppati insieme. I dati sono organizzati graficamente in un formato tabulare in cui ogni riga di box colorati rappresenta i livelli di trascritto di un gene misurati in vari istanti temporali ed ogni colonna rappresenta i livelli di trascritto di tutti i geni in un dato istante temporale secondo quanto viene rilevato da un microarray.

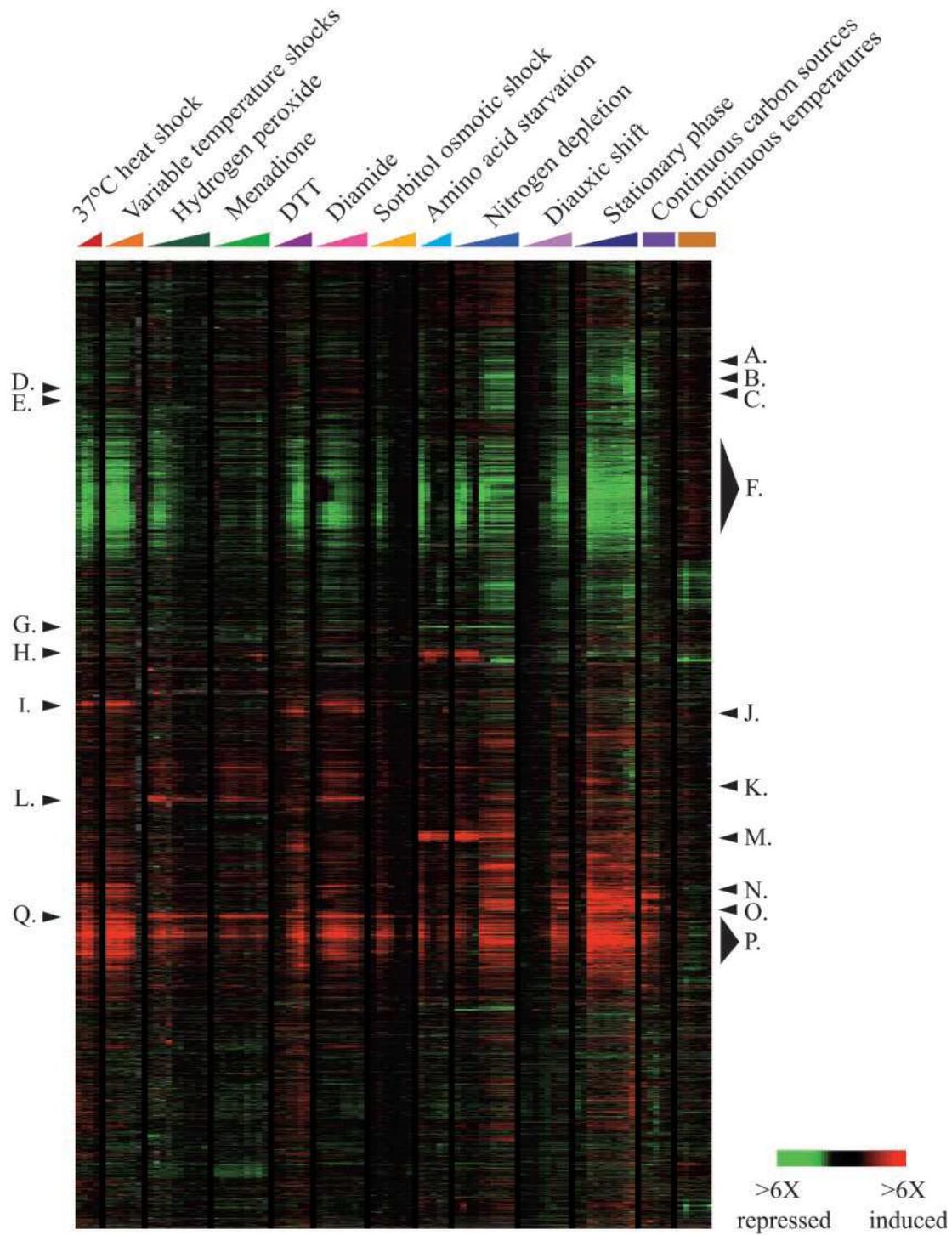


Figura 2.2 Livelli di espressione del genoma del *S. Cerevisiae* misurati in 94 microarray in risposta a differenti stimoli. L'intero set di geni è clusterizzato in base ai profili di espressione simili. Sulle righe sono presenti i geni, sulle colonne i diversi esperimenti. Il colore rosso indica una sovraespressione, il verde una sottoespressione.

L'analisi ha rivelato che un elevato numero di geni risponde con un andamento stereotipato definito "Environmental stress response" (ESR), costituito da un rapido picco di espressione (nell'ordine della decina di minuti) seguito da un ritorno al livello pre-stimolo. Sono stati individuati due grandi cluster, uno di geni repressi (il gruppo indicato con F. nella Figura 2.2) e uno di geni indotti (P. in Figura 2.2).

I due cluster mostrano profili identici ma reciproci. Nella figura 2.3 vengono riportati i profili medi di espressione nel caso specifico di due esperimenti consistenti nella transizione termica rispettivamente da 25°C a 37°C (a) e da 29°C a 33°C (b).

Il profilo dei geni indotti è indicato dalla curva rossa mentre quello dei geni repressi dalla curva verde. Si vede chiaramente come i profili siano identici nell'andamento temporale nonostante siano reciproci.

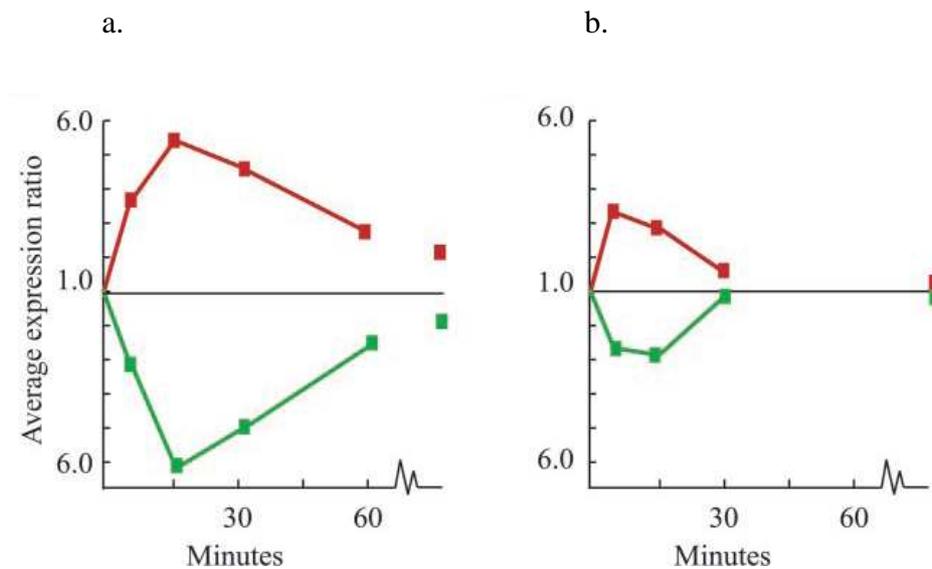


Figura 2.3 Profili medi di espressione di circa 1000 geni in risposta a shock termico da 25°C a 37°C (a) e da 29°C a 33°C (b). L'espressione media dei geni indotti è rappresentata dalla curva rossa, quella dei geni repressi è rappresentata dalla curva verde.

2.2. Reti di regolazione.

Abbiamo visto che mentre il DNA è lo stesso per tutte le cellule e per tutta la vita dell'organismo, non tutti i geni sono espressi in tutte le cellule e in ogni momento. Questo significa che esistono dei meccanismi che permettono di regolare l'espressione dei geni in base alle necessità della specifica cellula in ogni istante di tempo.

I meccanismi di regolazione sono molto complessi e, in un organismo eucariota, possono operare a ben 6 diversi livelli:

1. trascrizione;
2. processamento del trascritto (splicing, 5' capping, polyA Tail, ...);
3. esportazione degli mRNA completi al citoplasma;
4. traduzione;
5. degradazione dell'mRNA;
6. attivazione o disattivazione delle proteine.

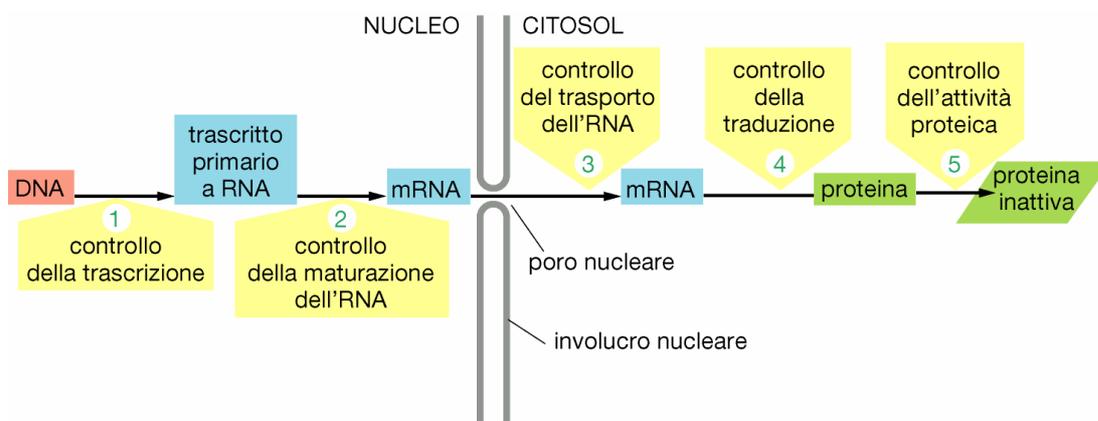


Figura 2.4 Possibili punti di controllo dell'espressione genica in un organismo eucariote.

Le reti di regolazione genica o GNR (dall'inglese Genetic Regulation Networks) descrivono le complesse interazioni che influenzano l'espressione genica e, conseguentemente, il comportamento cellulare.

Una rete è un grafo definito in matematica come una coppia $G=(V,E)$ in cui V è un insieme finito ed E è una famiglia di coppie di elementi di V . Gli elementi di V sono detti nodi o vertici mentre le coppie di E sono detti archi o lati. Esistono diverse tipologie di reti in base alla tipologia di archi utilizzati, si parla infatti di : reti orientate o non orientate (se gli archi hanno o

meno una direzione fissata) e reti pesate o non pesate (se ad ogni arco viene o meno assegnato un valore/peso).

Si parla di reti in molteplici contesti differenti quali matematica, fisica, biologia, sociologia ed economia. Tutte le reti sono simili come struttura in quanto sono sempre costituite da nodi e rami e possono essere analizzate con gli stessi criteri, ma differiscono per il significato che nodi e rami assumono nei diversi casi.

Negli ultimi anni si sta manifestando un crescente interesse per quanto riguarda lo studio delle interazioni esistenti tra le varie molecole presenti nelle cellule. L'obiettivo che cercano di raggiungere numerosi studiosi, utilizzando differenti tecniche di approccio, è quello di ricostruire i meccanismi di regolazione che coinvolgono geni, proteine e metaboliti a diversi livelli, tentando di ricostruire le "biochemical networks". In base al livello di dettaglio sul quale si lavora, le reti biochimiche possono essere suddivise nelle seguenti tre categorie:

- 1) reti geniche: rappresentano le relazioni che si instaurano tra geni andando a vedere in che modo il livello di espressione di un gene influisce sul livello di espressione degli altri;
- 2) reti proteiche: rappresentano le interazioni che avvengono tra le proteine quali la formazione di complessi e le modifiche proteiche causate dagli enzimi di signaling (in questo caso si parla anche di reti di signaling);
- 3) reti metaboliche: rappresentano le reazioni chimiche che avvengono tra i metaboliti.

Ognuno di questo tipo di reti è una semplificazione del sistema cellulare completo che può essere definito come "global biochemical network" il quale include tutti e tre i tipi di molecole (metaboliti, proteine e mRNA). Un' utile interpretazione grafica della struttura a livelli delle reti biochimiche è quella spesso utilizzata da Mendes e presentata in Figura 2.5.

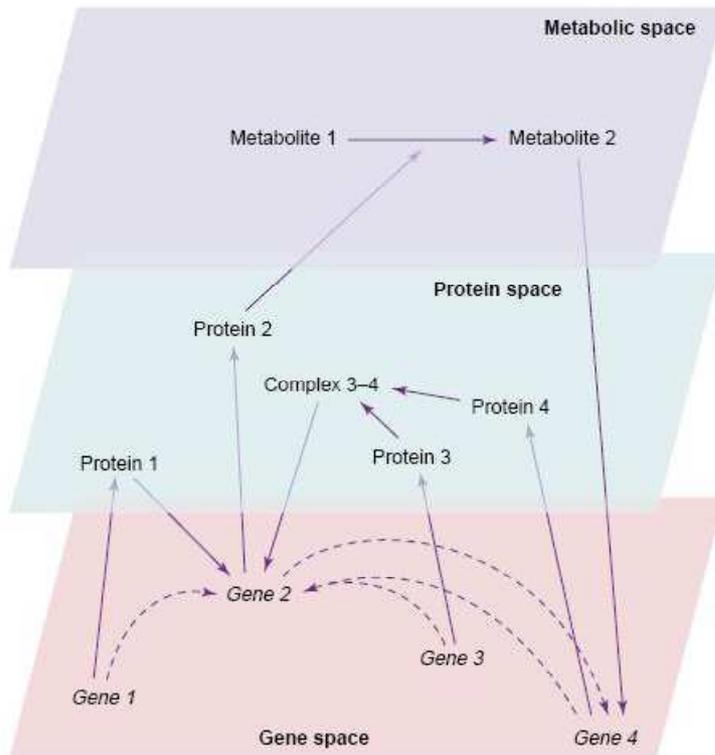


Figura 2.5 Esempio di rete biochimica organizzata a tre livelli: uno per i geni (o l'mRNA a seconda dell'interpretazione), uno per le proteine e uno per i metaboliti. Le frecce con linea continua indicano le interazioni tra i vari elementi, quelle tratteggiate costituiscono le proiezioni sul piano dei geni delle varie interazioni.

Poiché le tecnologie attuali non permettono di monitorare l'espressione delle proteine, come invece i microarray fanno per i trascritti, si è cercato innanzitutto di concentrarsi sull'utilizzo delle reti geniche. Se ci si limita a considerare il piano dei geni, quello che si fa è una proiezione di tutte le interazioni che coinvolgono geni, proteine e metaboliti su un unico livello.

In Figura 2.5 si possono vedere tre possibili differenti significati delle interazioni gene-gene: la regolazione del gene 2 avviene ad opera della proteina prodotta dal gene 1; la regolazione sempre del gene 2 ma effettuata da un complesso formato dalle proteine 3 e 4 prodotte rispettivamente dai geni 3 e 4; la regolazione del gene 4 da parte del metabolita 2, il quale viene prodotto grazie alla presenza della proteina 2.

Spesso le reti di regolazione vengono rappresentate graficamente utilizzando cerchi o punti per identificare i nodi e linee o frecce per identificare gli archi. La direzione delle interazioni, se presente, è indicata dall'orientazione degli archi stessi, che quindi sono rappresentati con una freccia. I pesi degli archi tuttavia dipendono dal modello che ipotizziamo per la rete ma in genere indicano l'intensità del controllo. Risulta evidente che tutti gli archi con

peso nullo sono regolazioni inesistenti e viceversa, quindi è possibile non rappresentarli graficamente.

Risulta evidente che l'elemento di maggior interesse della rete di regolazione genica così definita sono gli archi, ossia le regolazioni, che intercorrono tra due geni: essi sono individuati da coppie ordinate, se la rete è direzionale, (i,j) dove per convenzione si antepone il nodo controllato al nodo che controlla e quindi questo corrisponde alla freccia che va da j a i . A questo è associato un peso w_{ij} che mantiene nei pedici la stessa forma (si veda per maggiore chiarezza la Figura 2.6).

	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	w_{31}	w_{32}	0	0	0	0
4	0	0	w_{43}	0	w_{45}	0
5	0	0	0	w_{54}	0	0
6	0	0	0	w_{64}	0	0

Figura 2.6 Schema esemplificativo di una matrice di pesatura che rappresenta le connessioni tra i nodi da 1 a 6.

In tale modo la rete può essere riassunta in una matrice quadrata, che risulta una forma più adatta all'elaborazione al calcolatore, con un numero di righe e colonne pari al numero dei nodi che fungono da regolatori. Chiaramente sulle righe avremo i nodi in entrata mentre sulle colonne i nodi da cui gli archi si dipartono. Come detto in precedenza se l'arco non è presente ha peso nullo. Per quanto riguarda il segno del peso dell'arco, questo va a indicare il tipo di regolazione: un segno positivo indicherà un "up-regolazione" ovvero il controllo viene effettuato con proporzionalità diretta: se l'espressione di j cresce allora anche quella di i cresce, se cala allora anche quella di i cala. Viceversa, se il segno è negativo questa viene detta "down-regolazione" e sta a significare che la regolazione avviene con proporzionalità inversa: se l'espressione di j aumenta allora quella di i cala se quella di j cala allora quella di i aumenta. In letteratura sono stati proposti diversi metodi di reverse engineering per inferire sulle GNR a partire dai dati dinamici di espressione genica di microarrays dei quali verrà dato un breve accenno nel capitolo 3.

2.2.1 Rete trascrizionale di *Saccharomyces cerevisiae* e FFL

Nel corso degli ultimi anni le reti di regolazione genica sono state ampiamente analizzate ed una porzione della rete trascrizionale del lievito che comprende 690 geni è disponibile nel sito <http://www.weizmann.ac.il/mcb/UriAlon>. Svariati studi condotti dal gruppo di Uri Alon [5] [6] [7] [8] del Weizman Institute of Science, hanno permesso l'individuazione, anche all'interno della suddetta rete, di alcuni semplici pattern di regolazione, i cosiddetti "network motifs", di seguito chiamati motivi. I motivi di regolazione sono definiti come pattern che compaiono in una rete reale con una frequenza maggiore di quella con cui compaiono in una rete random, intesa come rete che ha un numero di nodi e di archi simile alle reti reali, ma connessioni completamente random.

Una volta individuati, i motivi di regolazione sono stati formalizzati attraverso semplici modelli, ed il loro comportamento è stato esaminato prima con tecniche di simulazione e poi validato sperimentalmente. Di recente i motivi sono stati sistematicamente definiti analizzando la rete di trascrizione di *E. Coli* [5], e poi ritrovati in altre reti di trascrizione di organismi animali e vegetali, ad esempio del lievito [9] come pure in reti che descrivono le modifiche post-traduzionali delle proteine e le interazioni tra neuroni.

I più importanti di questi pattern sono i seguenti tre (illustrati in Figura 2.7):

1) feed-forward loop (FFL): un fattore di trascrizione X regola un altro fattore di trascrizione Y ed entrambi collaborano alla regolazione di uno o più geni Z ;

2) single input module (SIM): un solo fattore di trascrizione X regola più geni Z e spesso si ha anche una autoregolazione di X . Generalmente tutte le regolazioni hanno lo stesso segno;

3) dense overlapping regulons (DOR): diversi geni Z_1, \dots, Z_m sono regolati ognuno da una diversa combinazione dei fattori di trascrizione X_1, \dots, X_n .

Si ritiene che questi pattern ricorrenti abbiano specifici ruoli funzionali all'interno della rete, possono quindi essere sfruttati per comprendere le dinamiche della rete in termini di componenti computazionali elementari mediante l'utilizzando di semplici modelli matematici. All'interno delle reti reali, però, questi motivi non sono isolati ma sono collegati tra loro, quindi sarà importante capire in quale misura sono preservate le loro caratteristiche, cioè capire se studiare il comportamento di motivi isolati permette di capire il funzionamento della rete globale.

Come vedremo in questo capitolo la nostra attenzione sarà rivolta al feed-forward loop, ossia al caso in cui un fattore di trascrizione X_1 regola un secondo fattore di trascrizione X_2 ed assieme vanno a regolare la trascrizione di X_3 . Quindi la regolazione tra X_1 e X_3 avviene con due meccanismi, uno diretto e l'altro mediato da X_2 . Nel FFL vi sono due segnali di input, due

induttori, S_{x_1} e S_{x_2} , che sono delle piccole molecole in grado di attivare o inibire l'attività trascrizionale di X_1 e X_2 . Dato che ciascuno dei tre archi ($X_1 \rightarrow X_2$, $X_2 \rightarrow X_3$, $X_1 \rightarrow X_3$) può corrispondere ad una attivazione o ad una repressione, ci sono $2^3 = 8$ possibili varianti dello stesso motivo. Queste otto varianti possono essere classificate in due gruppi: coerenti quando le due vie di regolazione, diretta ($X_1 \rightarrow X_3$) e indiretta ($X_1 \rightarrow X_2$, $X_2 \rightarrow X_3$) hanno lo stesso effetto, sia esso di attivazione o di repressione, incoerenti nel caso contrario, cioè quando le due vie hanno un effetto antagonista.

Il feed-forward loop può agire come uno switch sensibile a perturbazioni sia persistenti che transitorie. Ha inoltre la capacità di fornire il controllo temporale di un processo dato che l'espressione del gene target può dipendere dall'accumulazione di adeguati livelli di espressione e quindi di concentrazione degli altri due regolatori. In particolar modo il feed-forward loop (nella sua configurazione incoerente) è in grado di velocizzare la risposta del gene target in seguito ad uno stimolo a gradino, causando quindi un picco di espressione iniziale seguito dal ritorno ad un livello basale. Risulta quindi ragionevole pensare di adottare un motivo di tipo feed-forward loop per l'interpretazione del tipico profilo di risposta allo stress di cui abbiamo parlato nell'introduzione.

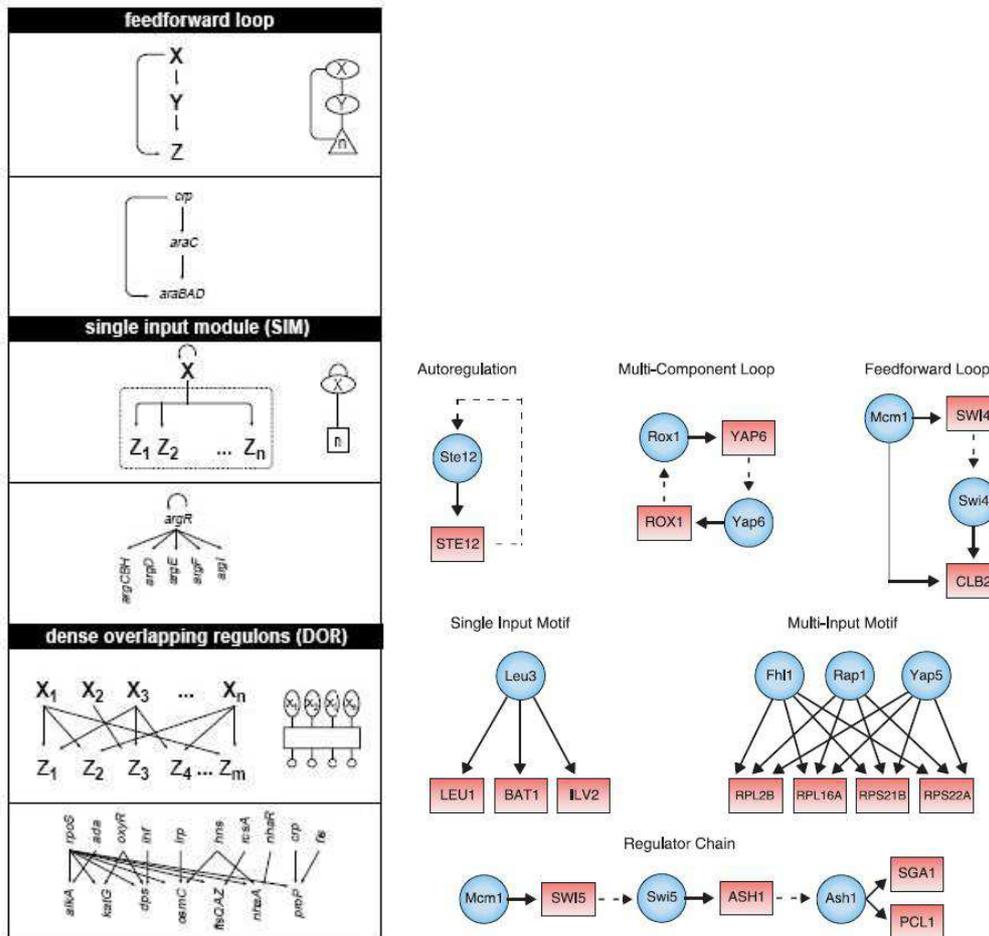


Figura 2.7 Rappresentazione grafica dei motifs caratteristici dell'Escherichia Coli (23) (immagine a sinistra) e del Saccharomyces Cerevisiae (26) (immagine in alto). Si può notare che i motifs ricorrenti nei due casi sono simili.

2.3. Associazione dei dati

Nei paragrafi 2.1.2 e 2.2.1 sono stati descritti i dati e la rete di regolazione del lievito che sono stati utilizzati per il presente lavoro di tesi. Il lavoro è iniziato quindi con l'associazione dei dati di espressione genica misurati da Gasch [1] in risposta a differenti stimoli, ai geni presenti nella rete di regolazione del S. Cerevisiae individuata da Alon [7].

Tra i motivi più ricorrenti che sono stati individuati all'interno delle reti, di cui abbiamo fatto cenno nel precedente paragrafo, la nostra attenzione si è rivolta in particolare al feed-forward loop (FFL). Esso prevede un fattore di trascrizione X_1 che regola un secondo fattore di trascrizione X_2 il quale, a sua volta, regola, assieme ad X_1 , la produzione della proteina X_3 . Nel FFL vi possono essere due segnali di input, S_{x1} e S_{x2} , che vanno ad attivare o inibire l'attività trascrizionale di X_1 e X_2 . Dalla letteratura [6] è noto che, in alcune sue configurazioni, il FFL è

in grado di velocizzare la risposta di X_3 in presenza di stimoli S_{x1} e S_{x2} e risulta quindi adatto a descrivere i pattern tipici della risposta allo stress caratterizzati da un picco iniziale e il successivo ritorno ad un livello basale. Se X_3 ha il tipico andamento ESR, il particolare profilo rilevato può essere causato dal meccanismo per cui i due segnali di input, che, secondo questa spiegazione, sarebbero interpretabili come dei segnali dovuti allo stress, innescano il loop nel quale i due fattori di trascrizione X_1 e X_2 regolano la trascrizione di X_3 per rispondere allo stress.

Nella rete trascrizionale di lievito, disponibile al sito <http://www.weizmann.ac.il/mcb/UriAlon>, sono stati individuati i feed-forward loop, ottenendo così 76 triplette di geni. Questi geni sono stati quindi associati ai rispettivi profili di espressione misurati da Gasch, disponibili al sito http://genome-www.stanford.edu/yeast_stress.

Dei 13 stimoli presenti nel database, sono stati considerati i 5 analizzati anche da Altafini [3], quindi ad ogni gene sono stati associati i profili temporali di espressione che derivano da 5 esperimenti differenti: due shock termici ($25^\circ \rightarrow 37^\circ$ e $29^\circ \rightarrow 33^\circ$), esposizione a perossido di idrogeno, esposizione all'agente sulfidril-ossidasi e shock iperosmotico.

Tra i 76 FFL individuati, ne sono stati selezionati 12 che soddisfano le seguenti caratteristiche: i geni target (X_3) hanno l'andamento che caratterizza la risposta allo stress (picco iniziale seguito dal ritorno al livello basale) e i regolatori (X_1 e X_2) sono differenzialmente espressi (è necessario che il loro profilo non sia piatto per poter vedere l'effetto della loro regolazione sul gene target). Per comodità abbiamo associato ad ogni FFL della rete un numero, perciò d'ora in poi faremo riferimento ai motivi utilizzando il numero che li individua. Per 2 dei 12 FFL analizzati, i geni target erano stati clusterizzati anche in [1] tra i geni caratterizzati da una evidente risposta allo stress. Si tratta dei seguenti FFL:

1. FFL numero 22: YER040W (X_1) – YKR034W (X_2) – YLR142W (X_3)
2. FFL numero 25: YER040W (X_1) – YKR034W (X_2) – YGR019W (X_3)

In figura 2.8 sono riportati i plot dei dati relativi ai tre geni che formano il FFL 22 e il FFL 25.

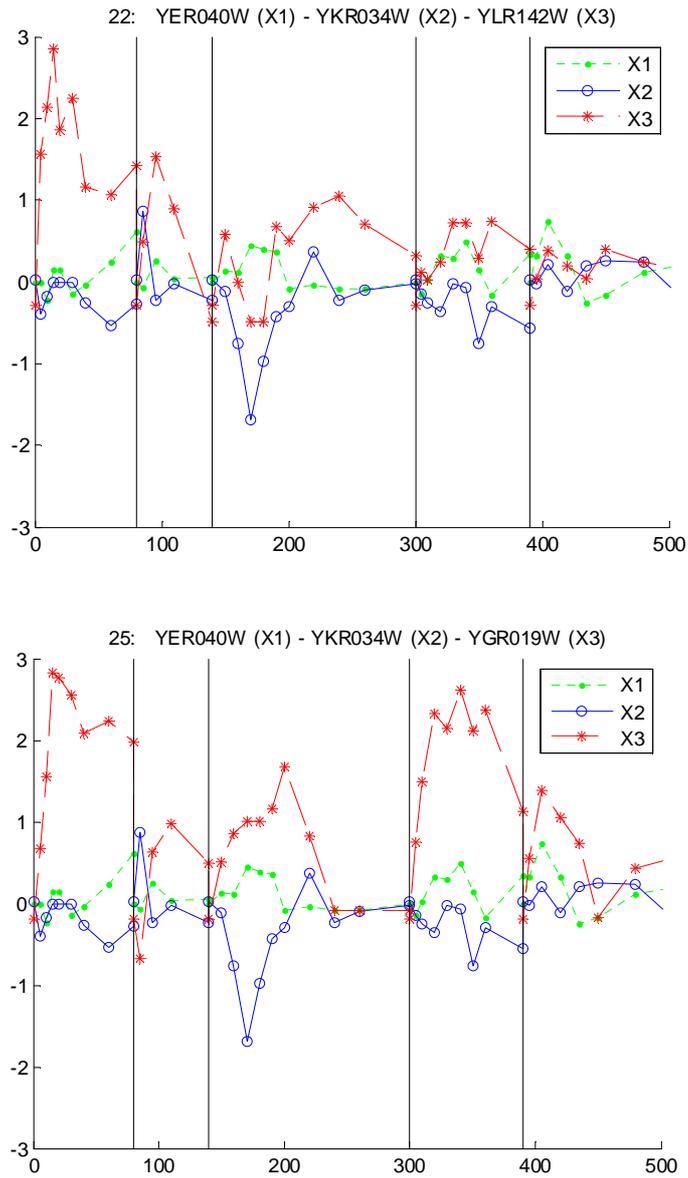
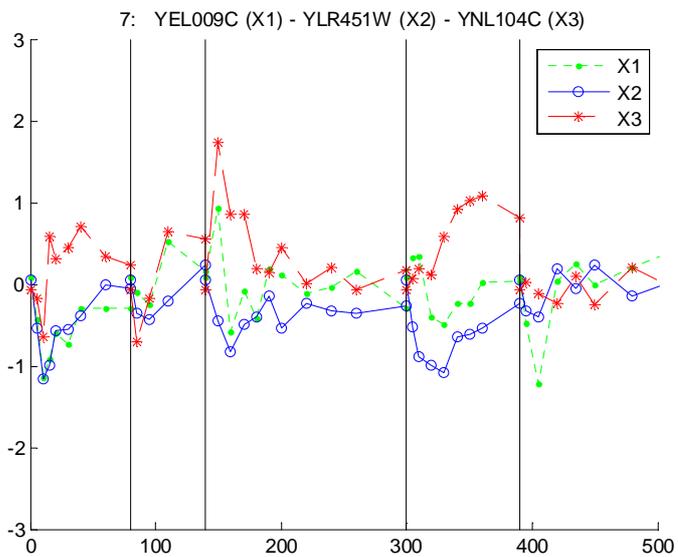
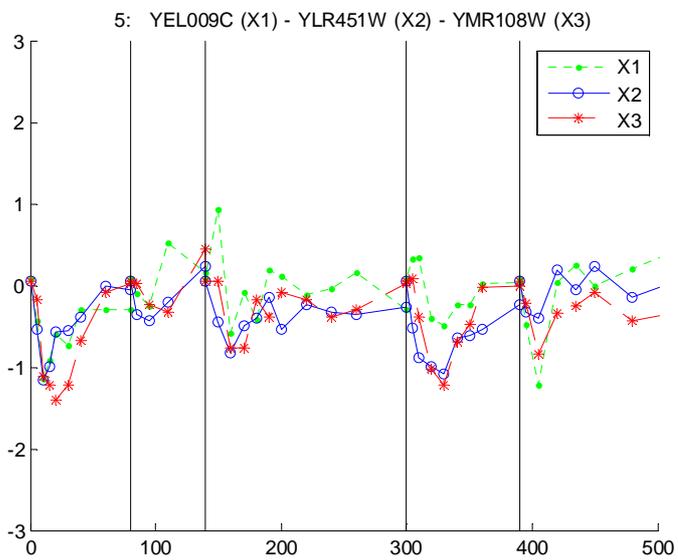
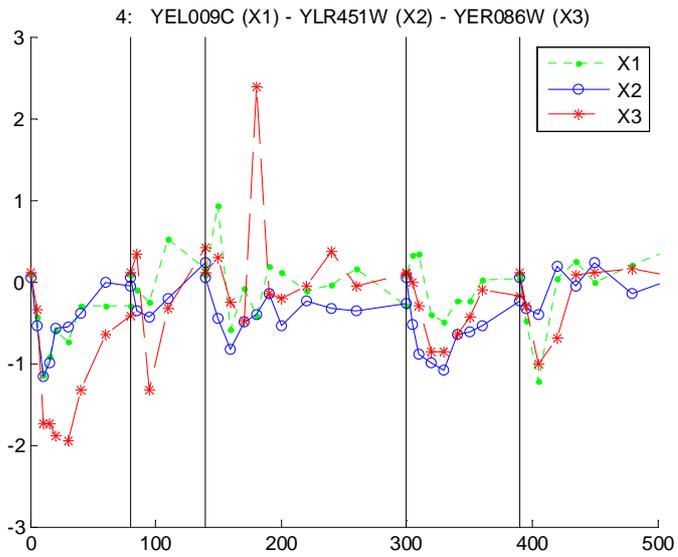
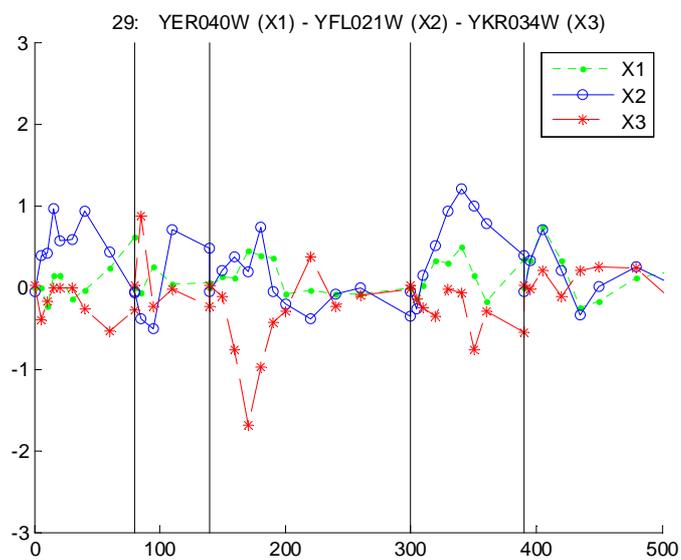
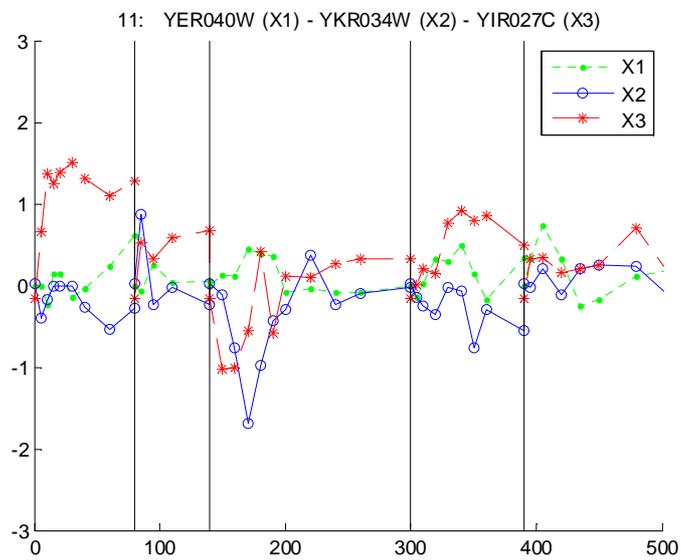
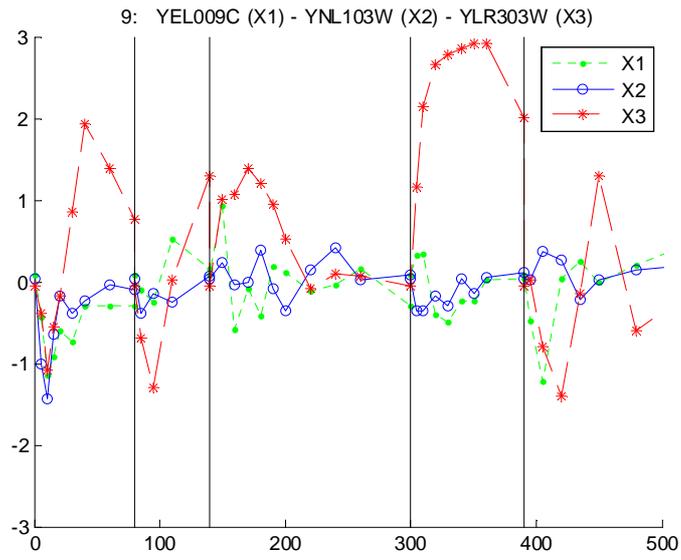
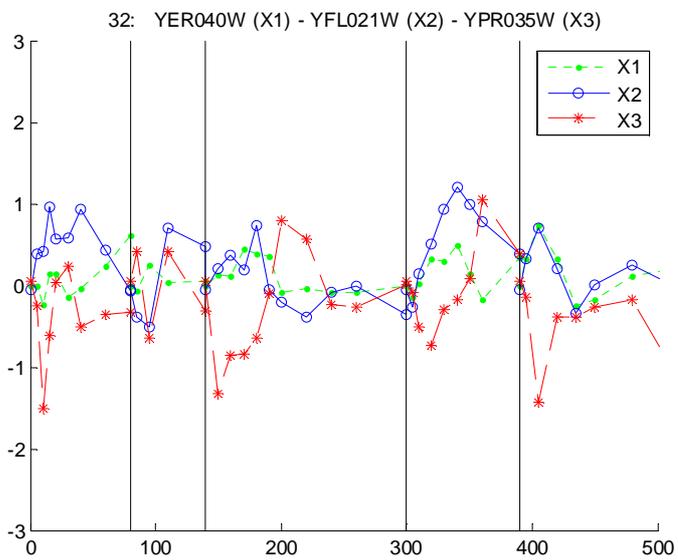
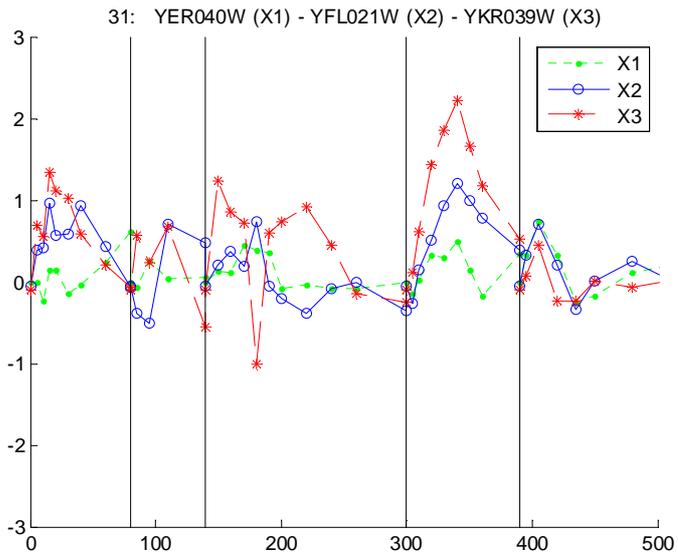


Figura 2.8 Plot dei dati relativi ai FFL 22 (sopra) e 25 (sotto).

In figura 2.9 sono riportati i plot dei dati di espressione relativi agli altri 10 FFL.







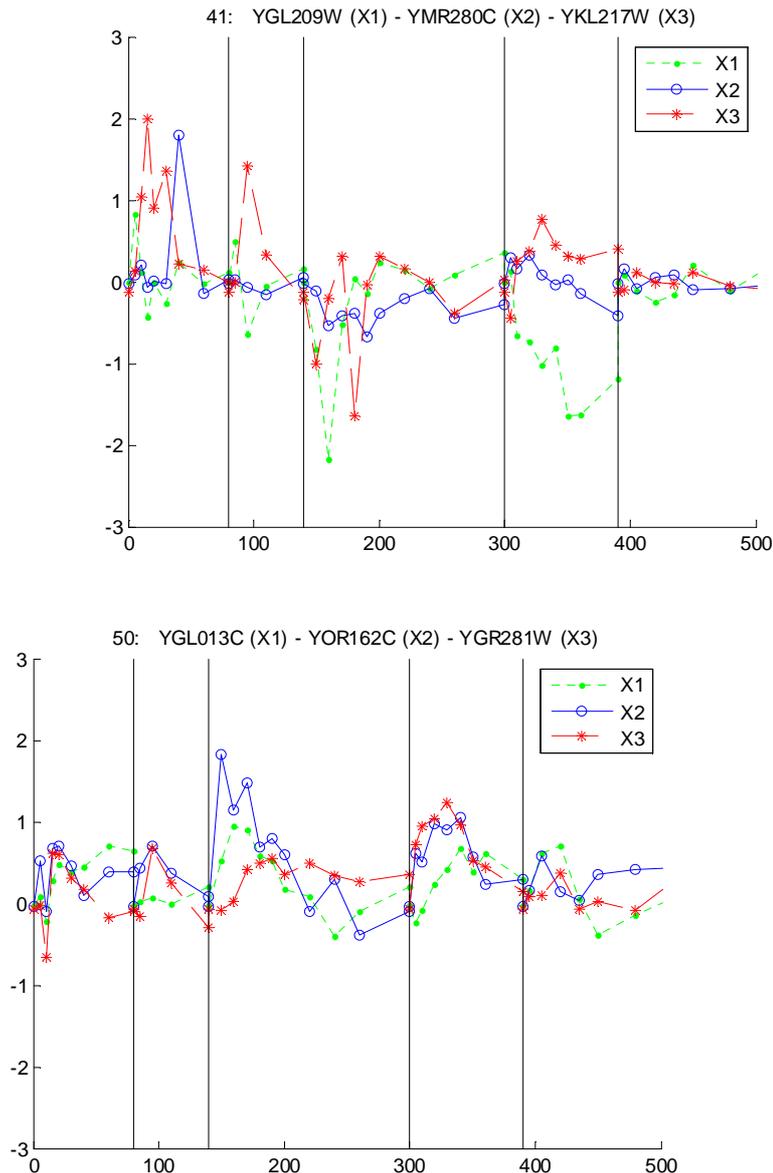


Figura 2.9 Plot dei dati relativi ai 10 FFL scelti.

Nei prossimi capitoli cercheremo di stimare numericamente il valore dei parametri incogniti dei due modelli (identificazione parametrica) a partire dai dati sperimentali che abbiamo qui descritti.

Confronteremo quindi la capacità dei due modelli di predire i profili con andamento ESR guardando non solo al fit, ovvero a come essi predicono i dati, ma anche alla precisione delle stime dei parametri ottenuti. Di conseguenza cercheremo di capire se il modello proposto in questo lavoro di tesi può essere considerato una valida alternativa al modello proposto dal gruppo di Altafini.

Capitolo 3

I modelli

3.1. Reverse engineering

Il termine reverse engineering indica l'insieme dei metodi, tipici dell'ingegneria dei controlli, utili a ricostruire la complessa rete di regolazione e controllo dall'output dinamico del sistema osservato. Data la complessità del sistema analizzato, gli approcci di reverse engineering per lo studio dell'interazione tra geni, proteine ed altri metaboliti è in genere limitato a piccole parti della rete di regolazione. Spesso, in questi casi, la rete di regolazione che si vuole studiare è già parzialmente nota, in alcuni casi a livello di dettaglio biochimico, e lo studio, così come la scelta dei geni e delle proteine da monitorare è "hypothesis driven", si basa cioè sulla conoscenza a priori e sulle ipotesi scientifiche che si vogliono verificare.

L'avvento delle tecnologie high-throughput ha dato la possibilità di passare da un approccio "hypothesis driven" ad un approccio "question driven" in cui, cioè, viene analizzato il sistema nel suo complesso invece di scegliere a priori le molecole da monitorare. In effetti oggi si ha a disposizione l'informazione relativa all'intera sequenza del genoma e tecnologie high-throughput come i microarray e la spettrometria di massa forniscono la possibilità di monitorare l'intero trascrittoma e traduttoma di un organismo\ tessuto\ cellula, in un dato istante, in una determinata condizione fisiologica.

È pertanto diventato di fondamentale importanza sviluppare modelli di reverse engineering adeguati a trattare la gran mole di trascritti monitorati in quanto questo tipo di studi permette di approfondire la conoscenza del sistema di regolazione e di interazione tra sottoreti diverse; inoltre aiuta ad identificare le sottoporzioni della rete di maggior interesse, su cui è possibile concentrarsi a posteriori e ad un più alto livello di dettaglio.

Le misure ricavate da un esperimento con microarray possono essere rappresentate con una matrice di valori di espressione genica le cui righe corrispondono ciascuna ad uno specifico trascritto monitorato e le colonne ai vari array. In genere è possibile monitorare l'intero genoma

(migliaia o decine di migliaia di trascritti) a fronte di qualche decina o centinaia di array (sia per motivi tecnici ed etici legati alla raccolta dei campioni, che per ragioni di costo). Pertanto i modelli di regolazione applicati a questo tipo di dati non possono essere molto complessi e includere un elevato numero di parametri. Inoltre, attraverso le tecnologie high-throughput non è possibile monitorare tutti i diversi tipi di molecole che partecipano al processo di regolazione e controllo pertanto la regolazione della trascrizione e della traduzione può essere ad esempio modellizzata come in Figura 3.1 (riquadro a) dove viene messo in evidenza il controllo di tipo feedback delle proteine sulla trascrizione mentre viene omesso il controllo effettuato a livello di RNA e altre molecole. Poiché la spettrometria di massa non permette di monitorare direttamente le proteine ma frammenti peptidici, attraverso i quali si deve risalire alla proteina di origine attraverso complesse tecniche di analisi, le misure high-throughput per la ricostruzione delle reti di regolazione sono spesso limitate alle misure del trascrittoma tramite microarrays. Quindi i modelli di regolazione sono di fatto ulteriormente semplificati come indicato in Figura 3.1 (riquadro b) in cui il livello di espressione dell'RNA è considerato come un'approssimazione del livello di espressione proteica delle proteine codificate dall'RNA stesso. Una rete di regolazione basata su quest'ultimo modello è detta "rete di regolazione genica" e può essere rappresentata come un grafo diretto in cui i nodi rappresentano i geni e le proteine da essi codificate mentre gli archi orientati rappresentano l'azione regolatrice tra nodi diversi.

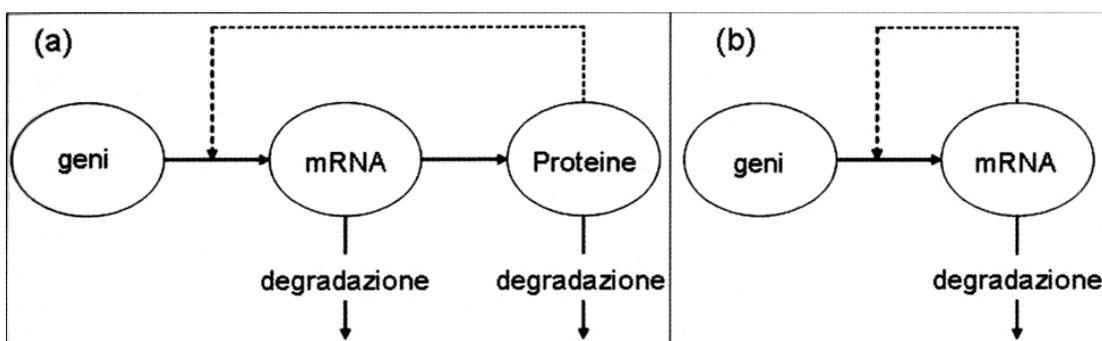


Figura 3.1 Modello di regolazione genica con controllo di tipo feedback della trascrizione da parte delle proteine (riquadro a) e dell'RNA messaggero (riquadro b), considerato come un'approssimazione della concentrazione proteica. Nel modello è stato omesso il controllo effettuato a livello di RNA e altre molecole.

Negli ultimi anni in letteratura sono stati proposti diversi metodi di reverse engineering per inferire sulla rete di regolazione dei dati dinamici di espressione genica di microarrays. Tali metodi si possono suddividere in tre grandi categorie: 1) metodi basati su misure di correlazione pair-wise tra coppie di geni, 2) metodi basati su modelli, 3) metodi basati sull'integrazione dei dati.

I primi confrontano profili di espressione di tutte le coppie di geni alla ricerca di possibili relazioni causa-effetto messe in luce dai profili stessi. I secondi definiscono un modello di regolazione e utilizzano i dati per identificare i parametri del modello e/o massimizzare una funzione punteggio che permette di scegliere tra configurazioni alternative del modello. I terzi combinano dati sperimentali con informazioni provenienti da data-base o da altri esperimenti nel tentativo di superare i limiti intrinseci di un'analisi condotta solo sui dati di espressione.

Metodi basati su modello.

I metodi basati su modello si possono suddividere in tre gruppi principali: modelli Booleani, modelli basati su equazioni, differenziali e non, e modelli Bayesiani.

Di seguito verranno brevemente descritti solo i modelli basati su equazioni differenziali dato che il presente lavoro di tesi andrà a concentrarsi su questi.

Metodi basati sulle equazioni differenziali.

I metodi basati sulle equazioni differenziali assumono la derivata dell'espressione genica di un gene x_i come funzione dell'espressione degli altri geni e di eventuali input esterni. Tale funzione può essere lineare o non lineare e più o meno complessa a seconda delle ipotesi sul modello.

Un modello semplice e abbastanza generale della regolazione assume il controllo del gene i come funzione di una combinazione lineare degli altri geni ($\sum w_{ij}x_j$) e può essere rappresentato, al tempo t e per il gene j , dalla seguente equazione differenziale:

$$\frac{dx_i(t)}{dt} = R_i \cdot f \left(\sum_{j=1}^N w_{ij} \cdot x_j(t) + \sum_{k=1}^K v_{ik} \cdot u_k(t) + B_i \right) - g \left(\sum_{j=1}^N h_{ij} \cdot x_j(t) \right) \quad (1)$$

dove

x_j è il valore di espressione osservato per il gene j ,

R_i è la costante di attivazione del gene i ,

u_k è la concentrazione dell'input esterno k ,

B_i è il livello di attivazione basale del gene i ,

w_{ij} e v_{ik} sono i parametri di controllo assunti indipendenti dal tempo. Sono positivi, negativi o nulli a seconda che il gene j o l'input k eserciti su i un controllo rispettivamente positivo, negativo o nullo,

f è la funzione di attivazione che può essere assunta ad esempio lineare o sigmoideale a seconda delle ipotesi sul modello,

g è la funzione di controllo della degradazione,

h_{ij} sono i parametri che controllano la degradazione (può essere pilotata da altri RNA o da altre proteine) del gene i .

I parametri w_{ij} , v_{ik} e h_{ij} sono incogniti e devono essere identificati risolvendo il sistema di equazioni differenziali a partire dai dati di espressione genica nei vari istanti temporali (i parametri significativamente diversi da zero identificano i regolatori del gene i e la loro efficacia nel regolare i).

$$\begin{array}{c} \text{time arrays} \\ \text{genes} \end{array} \begin{bmatrix} x_1(t_1) & \dots & x_1(t_M) \\ \dots & \dots & \dots \\ x_N(t_1) & \dots & x_N(t_M) \end{bmatrix}$$

Ed il sistema di equazioni per il gene 1 è:

$$\dot{x}_1(t_1) = R_i \cdot f(w_{11}x_1(t_1) + \dots + w_{1N}x_N(t_1)) - \lambda_1 x_1(t_1)$$

$$\dot{x}_1(t_2) = R_i \cdot f(w_{11}x_1(t_2) + \dots + w_{1N}x_N(t_2)) - \lambda_1 x_1(t_2)$$

....

$$\dot{x}_1(t_M) = R_i \cdot f(w_{11}x_1(t_M) + \dots + w_{1N}x_N(t_M)) - \lambda_1 x_1(t_M)$$

Per risolvere tale problema il numero di dati dovrebbe essere almeno pari al numero di parametri da identificare. Questa condizione difficilmente è soddisfatta dato che il numero di array disponibili (M) è generalmente minore del numero di geni analizzati (N) anche a valle di un

processo di selezione dei geni di interesse. Nel caso in cui, quindi, si verifichi che $M < N$, il sistema è indeterminato ed è necessario ricorrere a metodi euristici di ricerca della soluzione, come ad esempio diminuire artificialmente il numero di profili analizzati tramite Cluster Analysis o aumentare artificialmente il numero di array ricampionando una versione continua del pattern di espressione.

Altre soluzioni ipotizzano che la rete sia poco connessa (i dati finora a disposizione suggeriscono che ogni gene interagisca con un numero limitato di altri geni) e impongono dei vincoli all'equazione (1) per cui al massimo 10-12 pesi w_{ij} risultino diversi da zero per ogni gene i .

I modelli basati su equazioni differenziali fanno assunzioni esplicite sul modello di regolazione e pertanto sono più indicati per studiare quei sistemi di cui si conoscono i dettagli a livello della biochimica o comunque della dinamica del sistema. Presentano il vantaggio di considerare i dati in un range continuo di espressione e di poter rappresentare agevolmente feedback positivi e negativi nel modello di regolazione. Nel caso venga monitorato un numero limitato di variabili rispetto ai campioni a disposizione, questi modelli sono anche facilmente estendibili a modelli auto regressivi e quindi consentono di considerare ritardi gene-specifici nella trascrizione.

Tra gli aspetti negativi c'è la necessità di formulare un modello esplicito della regolazione (si noti però che un modello lineare può sempre essere interpretato come un'approssimazione di un modello complesso in situazioni di stato quasi stazionario) e il fatto che spesso il modello ha troppi parametri da identificare rispetto alla quantità di dati a disposizione come abbiamo appena visto.

L'uso di equazioni differenziali presuppone che le variabili in gioco varino in modo deterministico; tuttavia nelle reti di regolazione biologiche è possibile che due sistemi aventi le stesse condizioni iniziali terminino in due stati differenti a causa delle fluttuazioni nei tempi di trascrizione e traduzione. È possibile estendere i modelli basati su equazioni deterministiche a modelli stocastici in cui si tiene conto in termini probabilistici dell'evoluzione temporale dei fenomeni.

3.2 Costruzione dei modelli

L'interazione tra gli elementi costitutivi di un modulo di regolazione può essere modellata matematicamente mediante l'utilizzo di equazioni differenziali ordinarie lineari o non lineari la cui soluzione descrive la cinetica temporale del trascritto e delle proteine da esso codificate.

3.2.1 Modello “Feedback integrativo”

Nel lavoro di Altafini [3] il transiente tipico dell’espressione genica in risposta ad uno stress (ESR), viene descritto in modo qualitativo con un modello del tipo:

$$\dot{X}_i(t) = -F_{0i}(t) + R_i(t)$$

dove si indica con X_i l’espressione del gene i -esimo, la quale viene modulata da una equazione differenziale in cui è presente un termine di degradazione $F_{0i}(t)$ e uno di produzione $R_i(t)$. I termini di degradazione e produzione che sono stati presi in considerazione sono i seguenti:

$$F_{0i}(t) = \delta_{X_i} X_i(t)$$

$$R_i(t) = c + b_i u(t) - a_i P_i(t)$$

La degradazione viene quindi considerata proporzionale alla concentrazione stessa di mRNA. La produzione è invece composta da tre termini: un primo termine costante che serve a mantenere la quantità basale di mRNA, un secondo termine che descrive la variazione della produzione basale in seguito allo stimolo (descritta da un aumento a gradino $u(t)$). Il terzo termine rappresenta invece l’azione di feedback effettuata ad opera della proteina P_i associata al gene. L’andamento del prodotto genico viene descritto dalla seguente equazione differenziale:

$$\dot{P}_i(t) = r X_i(t) + \delta_{P_i} P_i(t)$$

Dove r rappresenta il tasso di traduzione e δ_{P_i} quello di degradazione della proteina. La proteina viene quindi prodotta in maniera proporzionale all’espressione del gene corrispondente e degradato proporzionalmente alla sua stessa concentrazione. Dato che i tempi di degradazione sono superiori alla finestra temporale che si prende in considerazione (un paio d’ore), il termine $\delta_{P_i} P_i(t)$ viene considerato trascurabile.

Il modello si presenta dunque nella seguente forma:

$$\begin{cases} \dot{X}_i = -\delta_i X_i + a_i P_i + b_i u + c \\ \dot{P}_i = r X_i \end{cases} \quad \begin{cases} X_i(0) = X_{ib} \\ P_i(0) = P_{ib} \end{cases}$$

Trattandosi di misure di spot array a due canali, per i dati di espressione genica le misure sono:

$$Z_i = \log_2 \left(\frac{B_{i,Cy5}}{B_{i,Cy3}} \right) = \log_2(B_{i,Cy5}) - \log_2(B_{i,Cy3}) = X_i - X_{ib}$$

In genere si considera la “log-ratio” poiché si è visto che la concentrazione dei trascritti risulta essere proporzionale al logaritmo del segnale, costituito dal rapporto tra le due intensità, piuttosto che al segnale stesso.

Passando al modello alle differenze rispetto alla concentrazione basale il modello diventa:

$$\begin{cases} \dot{X}_i = -\delta_{xi}(X_i - X_{ib}) - a_i(P_i - P_{ib}) + b_i u & X_i(0) = X_{ib} \\ \dot{P}_i = r(X_i - X_{ib}) & P_i(0) = P_{ib} \end{cases}$$

Con un cambio di variabili

$$x_i = X_i - X_{ib}, \quad p_i = P_i - P_{ib}$$

si ottiene:

$$\begin{cases} \dot{x}_i = -\delta_{xi}x_i - a_i p_i + b_i u & x_i(0) = 0 \\ \dot{p}_i = r x_i & p_i(0) = 0 \end{cases}$$

E le equazioni di misura sono:

$$Z_i = x_i$$

Oltre a porre $\delta = 0$, in [3] il valore di r è stato assunto uguale per tutti i geni ($r = 0.01$).

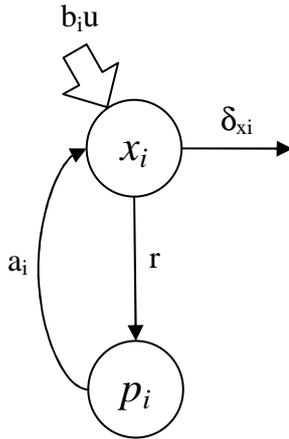


Figura 3.2 Schema del modello “feedback-integrativo”

3.2.2 Modello “Feed-forward loop”

Come spiegato nel capitolo 2, è stato formulato un modello alternativo a quello proposto da Altafini, considerando uno dei network motifs presenti nel lievito, ovvero il feed-forward loop. Si tratta di un loop che coinvolge 3 geni della stessa cellula: un gene X_3 regolato dai geni X_1 e X_2 , con X_1 che regola X_2 . Supponiamo dunque che X_3 sia il gene con la tipica risposta ESR che vogliamo modellizzare. È stato quindi considerato il seguente modello:

$$\begin{cases} \dot{X}_2 = S_2 + k_{21}X_1 - k_{02}X_2 & X_2(0) = X_{2b} \\ \dot{X}_3 = S_3 + k_{31}X_1 + k_{32}X_2 - k_{03}X_3 & X_3(0) = X_{3b} \end{cases}$$

Con equazioni di misura:

$$Z_1 = \log_2 \left(\frac{B_{1,cy5}}{B_{1,cy3}} \right) = \log_2(B_{1,cy5}) - \log_2(B_{1,cy3}) = X_1 - X_{1b}$$

$$Z_2 = \log_2 \left(\frac{B_{2,cy5}}{B_{2,cy3}} \right) = \log_2(B_{2,cy5}) - \log_2(B_{2,cy3}) = X_2 - X_{2b}$$

$$Z_3 = \log_2 \left(\frac{B_{3,cy5}}{B_{3,cy3}} \right) = \log_2(B_{3,cy5}) - \log_2(B_{3,cy3}) = X_3 - X_{3b}$$

Dove S_2 e S_3 rappresentano la produzione basale, k_{21} è il parametro di regolazione della traduzione di X_2 da parte di X_1 , k_{31} e k_{32} sono i parametri di regolazione della trascrizione di X_3 da parte di X_1 e X_2 rispettivamente e k_{02} e k_{03} sono i tassi di degradazione dei rispettivi mRNA.

In stato stazionario pre-stimolo si ha che:

$$S_2 + k_{21}X_{1b} - k_{02}X_{2b} = 0$$

$$S_3 + k_{31}X_{1b} + k_{32}X_{2b} - k_{03}X_{3b} = 0$$

Il modello alle differenze rispetto al basale si presenta nella seguente forma:

$$\begin{cases} \dot{X}_2 = k_{21}(X_1 - X_{1b}) - k_{02}(X_2 - X_{2b}) \\ \dot{X}_3 = k_{31}(X_1 - X_{1b}) + k_{32}(X_2 - X_{2b}) - k_{03}(X_3 - X_{3b}) \end{cases} \quad \begin{matrix} X_2(0) = X_{2b} \\ X_3(0) = X_{3b} \end{matrix}$$

Cambiando le variabili

$$x_1 = (X_1 - X_{1b}), \quad x_2 = (X_2 - X_{2b}), \quad x_3 = (X_3 - X_{3b})$$

si ottiene infine:

$$\begin{cases} \dot{x}_2 = k_{21}x_1 - k_{02}x_2 \\ \dot{x}_3 = k_{31}x_1 + k_{32}x_2 - k_{03}x_3 \end{cases} \quad \begin{matrix} x_2(0) = 0 \\ x_3(0) = 0 \end{matrix}$$

E le equazioni di misura sono:

$$Z_1 = x_1$$

$$Z_2 = x_2$$

$$Z_3 = x_3$$

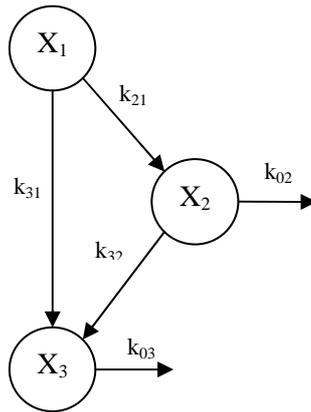


Figura 3.3 Schema del modello “feed-forward”.

3.3. Implementazione dei modelli

Nel paragrafo 3.1 abbiamo definito le equazioni differenziali che costituiscono i due modelli del sistema che stiamo indagando, ovvero la risposta allo stress.

Per completare ciascuno dei due modelli è necessario risolvere il problema cosiddetto della stima parametrica ovvero dare dei valori numerici ai parametri che vi compaiono: $[k_{31}, k_{32}, k_{03}]$ per il modello da noi proposto e $[a_i, b_i, \delta_i]$ per il modello proposto dal gruppo di Altafini.

Per fare ciò adottiamo lo stimatore ai minimi quadrati non lineari pesati.

Si tratta di una strategia risolutiva che sfrutta il modello, le misure e la descrizione statistica del secondo ordine dell'errore di misura .

Come abbiamo spiegato nel paragrafo 2.3, il nostro segnale misurato è la “log-ratio” (ovvero il logaritmo in base due del rapporto tra i segnali rilevati nei due canali)

$$LR = \log_2 \left(\frac{S_{Cy5}}{S_{Cy3}} \right) = \log_2(S_{Cy5}) - \log_2(S_{Cy3}) = C - C_{basale}$$

che corrisponde alla differenza tra la concentrazione di trascritto nell'istante campionato e la sua concentrazione basale.

Disponiamo di un numero limitato di campioni, affetti da errore di misura, il quale è assimilabile ad una variabile aleatoria di tipo gaussiano a campioni scorrelati con valore atteso zero e

varianza costante incognita. Per la precisione il numero di campioni che possediamo è diverso per ognuna delle cinque serie temporali che sono state considerate. Nella Tabella 1 sono riportati il numero di campioni e gli istanti temporali di ogni serie.

Serie temporale	Numero campioni	Istanti di campionamento										
		t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}
1	9	0	5	10	15	20	30	40	60	80		
2	5	0	5	15	30	60						
3	11	0	10	20	30	40	50	60	80	100	120	160
4	9	0	5	10	20	30	40	50	60	90		
5	8	0	5	15	30	45	60	90	120			

Tabella 1

Istanti di campionamento relativi alle cinque serie temporali.

Chiamiamo \mathbf{p} il vettore dei parametri da stimare. L'uscita predetta dal modello è una funzione del tempo e del vettore dei parametri e verrà indicata con

$$y(t) = g(t, \mathbf{p})$$

L'uscita misurata al tempo di campionamento t_k è la "log-ratio" ed è descrivibile come:

$$LR = z_k = y_k + v_k = g(t_k, \mathbf{p}) + v_k, \quad \text{con } k = 1, 2, \dots, N \quad (1)$$

dove y_k è l'uscita del modello al tempo t_k e v_k è l'errore di misura log-additivo che corrompe la k-esima misura z_k (N è il numero delle misure).

L'equazione (1) si può scrivere in forma vettoriale come

$$\mathbf{z} = \mathbf{y} + \mathbf{v} = G(\mathbf{p}) + \mathbf{v} \quad (2)$$

dove $G(\mathbf{p}) = [g(t_1, \mathbf{p}) \ g(t_2, \mathbf{p}) \ \dots \ g(t_N, \mathbf{p})]^T$.

La deviazione standard dell'errore di misura è costante ma incognita perciò verrà stimata a posteriori. In partenza la poniamo pari a 1. La matrice di covarianza del vettore \mathbf{v} è esprimibile nella forma:

$$\Sigma_v = E[\mathbf{v}\mathbf{v}^T] = \sigma^2 B$$

dove σ^2 è un opportuno scalare, incognito, e B è una matrice quadrata identica N -dimensionale.

Per il vettore \mathbf{p} dei parametri (non noto) si definisce l'errore di predizione come:

$$e(\mathbf{p}) = z - G(\mathbf{p})$$

ovvero come la differenza tra il vettore dei dati e quello delle predizioni del modello per quel determinato valore di \mathbf{p} . L'errore di predizione su una certa misura va ritenuto tanto più "grave" quanto più questa era attendibile.

Consideriamo lo scalare

$$\|z - G(\mathbf{p})\|_{\Sigma_v^{-1}}^2 = [z - G(\mathbf{p})]^T \Sigma_v^{-1} [z - G(\mathbf{p})] \quad (3)$$

che misura la distanza tra i dati z e la predizione del modello $G(\mathbf{p})$ pesata in maniera inversa alla matrice di covarianza dell'errore di misura. Lo stimatore ai minimi quadrati non lineari pesati determina il valore di \mathbf{p} minimizzando quella distanza:

$$\hat{\mathbf{p}} = \min_{\mathbf{p}} \|z - G(\mathbf{p})\|_{\Sigma_v^{-1}}^2 = \min_{\mathbf{p}} [z - G(\mathbf{p})]^T \Sigma_v^{-1} [z - G(\mathbf{p})]$$

Della matrice Σ_v è nota solo $B = I_N$ perciò si utilizza la dicitura "pesi relativi" e il problema è riformulato nel seguente modo:

$$\hat{\mathbf{p}} = \min_{\mathbf{p}} \|z - G(\mathbf{p})\|_{B^{-1}}^2 = \min_{\mathbf{p}} [z - G(\mathbf{p})]^T B^{-1} [z - G(\mathbf{p})] \quad (4)$$

Il fatto di dover usare, a causa della mancanza di σ^2 , i "pesi relativi" non influenza la stima del vettore dei parametri. Una volta risolta la (4), il valore di σ^2 è stimabile a posteriori dividendo il valore della funzione costo nel punto di minimo per il numero di gradi di libertà, definiti come la differenza tra il numero N dei dati usati per l'identificazione e il numero M dei parametri del modello.

$$\hat{\sigma}^2 = \frac{[z - G(\hat{\mathbf{p}})]^T B^{-1} [z - G(\hat{\mathbf{p}})]}{N - M} \quad (5)$$

Il calcolo di σ^2 è fondamentale perché ci permette di ottenere la precisione delle stime che infatti richiede la conoscenza di Σ_v .

Un criterio importante infatti al fine di giudicare la bontà di un modello è legato alla precisione con cui sono stimati i suoi parametri. I dati $\{z_k\}$ da cui si ottiene la stima del vettore dei parametri sono infatti incerti a causa dell'errore di misura pertanto anche il vettore delle stime dei parametri risulta incerto.

Definiamo l'errore di stima sul vettore dei parametri come:

$$\tilde{\mathbf{p}} = \mathbf{p} - \hat{\mathbf{p}} \quad (6)$$

Nella (6) \mathbf{p} è il valore “vero”, ma incognito, del vettore dei parametri. Nell'ipotesi che $E[\tilde{\mathbf{p}}]$ abbia media nulla, ovvero $E[\hat{\mathbf{p}}] = \mathbf{p}$ (non polarizzazione in media dello stimatore), si può misurare la precisione delle stime dalla matrice di covarianza dell'errore di stima

$$\text{cov}(\tilde{\mathbf{p}}) = E[\tilde{\mathbf{p}}\tilde{\mathbf{p}}^T] \quad (7)$$

che coincide con $\text{cov}(\hat{\mathbf{p}})$ e dalla cui diagonale è possibile ottenere la standard deviation della stima e perciò si possono costruire gli intervalli di confidenza della stima:

$$\hat{p}_i \pm SD(\hat{p}_i)$$

Nei casi pratici, ed è quello che si farà anche in questo lavoro di tesi, si preferisce riportare l'incertezza in termini relativi utilizzando in indice che prende il nome di “coefficiente di variazione”:

$$CV(\hat{p}_i) = 100 \times \frac{SD(\hat{p}_i)}{\hat{p}_i}$$

Nel caso di modelli non lineari, come sono i nostri, non è possibile calcolare analiticamente $\text{cov}(\hat{\mathbf{p}})$ ma si deve ricorrere ad un'approssimazione di questa

$$\text{cov}(\tilde{\mathbf{p}}) = \text{cov}(\hat{\mathbf{p}}) \cong (S^T \Sigma_v^{-1} S)^{-1}$$

dove S è la matrice “sensitività” così definita

$$S = \begin{bmatrix} \left. \frac{\partial g(t_1, p)}{\partial p_1} \right|_{p=\hat{p}} & \left. \frac{\partial g(t_1, p)}{\partial p_2} \right|_{p=\hat{p}} & \dots & \left. \frac{\partial g(t_1, p)}{\partial p_M} \right|_{p=\hat{p}} \\ \left. \frac{\partial g(t_2, p)}{\partial p_1} \right|_{p=\hat{p}} & \left. \frac{\partial g(t_2, p)}{\partial p_2} \right|_{p=\hat{p}} & \dots & \left. \frac{\partial g(t_2, p)}{\partial p_M} \right|_{p=\hat{p}} \\ \dots & \dots & \dots & \dots \\ \left. \frac{\partial g(t_{N-1}, p)}{\partial p_1} \right|_{p=\hat{p}} & \left. \frac{\partial g(t_{N-1}, p)}{\partial p_2} \right|_{p=\hat{p}} & \dots & \left. \frac{\partial g(t_{N-1}, p)}{\partial p_M} \right|_{p=\hat{p}} \\ \left. \frac{\partial g(t_N, p)}{\partial p_1} \right|_{p=\hat{p}} & \left. \frac{\partial g(t_N, p)}{\partial p_2} \right|_{p=\hat{p}} & \dots & \left. \frac{\partial g(t_N, p)}{\partial p_M} \right|_{p=\hat{p}} \end{bmatrix}$$

Il vettore dato dalla (4) non è ottenibile tramite calcoli analitici ma solo attraverso degli algoritmi numerici di tipo iterativo che normalmente hanno come punto assai delicato la scelta della stima iniziale del vettore dei parametri, che può influenzare la discesa verso un minimo locale piuttosto che globale.

Uno dei metodi più usati per la soluzione del problema (4) è quello di Gauss-Newton nelle sue versioni numericamente robuste (ad es. Levenberg-Marquardt).

Abbiamo implementato i modelli lavorando nell'ambiente di calcolo Matlab e, nello specifico, è stata utilizzata la versione Matlab R2009a.

Si è adoperata la funzione "lsqnonlin", inclusa nell'Optimization toolbox, che implementa lo stimatore ai minimi quadrati non lineari. La funzione sceglie, di default, l'algoritmo iterativo "Trust-region-reflective" basato sul metodo "Interior-reflective Newton".

La funzione lsqnonlin risolve dunque problemi della forma:

$$\hat{\mathbf{p}} = \min_p \sum_{i=1}^N F_i(\mathbf{p})^2$$

perciò richiede la definizione della funzione

$$F(\mathbf{p}) = \begin{bmatrix} F_1(\mathbf{p}) \\ F_2(\mathbf{p}) \\ \vdots \\ F_N(\mathbf{p}) \end{bmatrix}.$$

Abbiamo perciò implementato una function che fornisca in uscita il vettore che costituisce la funzione obiettivo $F(\mathbf{p})$ da minimizzare. Data la (4), la nostra funzione costo $F(\mathbf{p})$ ha la seguente forma:

$$F(\mathbf{p}) = \begin{bmatrix} \frac{z_1 - G_1(\mathbf{p})}{\sigma_1} \\ \frac{z_2 - G_2(\mathbf{p})}{\sigma_2} \\ \vdots \\ \frac{z_N - G_N(\mathbf{p})}{\sigma_N} \end{bmatrix}$$

Il vettore viene ricalcolato ad ogni passo della procedura di ottimizzazione (lsqnonlin).

La funzione lsqnonlin restituisce lo jacobiano della soluzione grazie al quale possiamo calcolare la matrice sensitività e quindi la covarianza delle stime infatti

$$\Sigma_{\hat{\mathbf{p}}} = \Sigma_{\hat{\mathbf{p}}} \cong (S^T \Sigma_v^{-1} S)^{-1} = (J^T J)^{-1}$$

Entrambi i modelli corrispondono ad un sistema di equazioni differenziali perciò abbiamo creato una funzione che, ad ogni istante di tempo, restituisca il valore delle derivate del sistema. Abbiamo usato il solutore “ODE45” che risolve equazioni differenziali non stiff usando metodi Runge-Kutta.

Per applicare il solutore è richiesta la creazione di una function in cui definire il modello. Nel caso del modello alternativo da noi proposto in questa tesi, basato sull'utilizzo di feed-forward loop, gli ingressi sono i profili dei due regolatori come descritto nel paragrafo 3.2.2 perciò abbiamo considerato come funzioni forzanti le interpolazioni lineari dei dati relativi ai due fattori di trascrizione.

Nel caso del modello di proposto da Altafini, invece, si considera come ingresso un gradino unitario.

Come abbiamo giudicato la bontà dell'identificazione parametrica?

Due importanti criteri sono l'analisi dei residui e la precisione delle stime.

Supponiamo di aver ottenuto una stima $\hat{\mathbf{p}}$ del vettore dei parametri. Il vettore dei residui è così definito:

$$\mathbf{r} = \mathbf{z} - G(\hat{\mathbf{p}})$$

In base all'equazione (2), il vettore costituisce, in un certo senso, la stima del vettore degli errori di misura \mathbf{v} . È logico quindi attendersi che un buon modello porti a residui che siano compatibili con la statistica dell'errore di misura. Nel nostro caso l'errore di misura ha media nulla e campioni scorrelati perciò l'esistenza di una media non nulla o la presenza di forti correlazioni nei residui (ad esempio una lunga sequenza di residui dello stesso segno) è una verosimile spia di una modellizzazione non corretta o di una ottimizzazione della funzione costo inefficace (ad esempio perché il metodo iterativo di discesa si è imbattuto in un minimo locale anziché essere pervenuto a quello globale). Noi considereremo la norma al quadrato del vettore dei residui normalizzati, cioè la somma quadratica dei residui pesati

$$WRSS = [\mathbf{z} - G(\hat{\mathbf{p}})]^T B^{-1} [\mathbf{z} - G(\hat{\mathbf{p}})]$$

che corrisponde al valore della funzione obiettivo calcolata per il vettore ottimo dei parametri. Si sceglie dunque il modello con minore indice $WRSS$. Per comodità, anziché valutare $WRSS$, abbiamo considerato la deviazione standard a posteriore che differisce solamente per un fattore di scala infatti:

$$sd_{post} = \frac{WRSS(\hat{\mathbf{p}})}{N - M}$$

L'altro criterio importante al fine di giudicare la bontà di un modello è legato alla precisione con cui sono stimati i suoi parametri. Guarderemo quindi il valore dei coefficienti di variazione associati ai parametri stimati. È chiaro che ottenere precisioni insufficienti delle stime (in relazione allo scopo dello studio) non è cosa accettabile e le cause possono essere molteplici. Se i residui sono accettabili, è probabilmente l'eccessiva complessità del modello per i dati a giocare un ruolo determinante. Altre possibili cause sono numero e dislocazione dei campioni e l'errore di misura associato ai dati.

Un miglioramento della precisione si potrebbe ottenere grazie ad un approccio bayesiano al problema della stima, che sfrutta l'informazione a priori disponibile in forma statistica sul vettore dei parametri.

Scelta tra modelli in competizione

Spesso ci si trova a dover scegliere il modello del sistema in una rosa di più modelli candidati. Se i modelli sono di ordine diverso non è possibile utilizzare la norma al quadrato del vettore dei residui normalizzati, cioè la somma quadratica dei residui pesati

$$WRSS = [z - G(\hat{p})]^T \Sigma_v^{-1} [z - G(\hat{p})] \quad (8)$$

per discriminare il migliore dei candidati. Infatti è facile rendersi conto che i modelli, al crescere del numero di parametri, descrivono meglio i dati. Tuttavia, il costo derivante dall'aumento del numero di parametri è un incremento a volte inaccettabile della precisione delle stime.

Il valore di WRSS viene calcolato dalla funzione `lsqnonlin` stessa che lo restituisce nel parametro "resnorm".

Nel confrontare la capacità di modelli di ordine diverso di descrivere i dati sperimentali bisogna quindi guardare non solo al cosiddetto fit, ovvero a come essi predicono i dati (cosa che migliora sempre all'aumentare dell'ordine del modello), ma anche alla precisione delle stime dei parametri ottenuti (che peggiora sempre).

Per trovare quale tra più modelli in competizione e di ordine diverso meglio bilanci le due esigenze contrastanti di fit e precisione, si utilizzano criteri basati sul principio di parsimonia. Ad esempio è disponibile il criterio di Akaike (Akaike, 1974) il quale suggerisce che, nel caso di errore di misura con varianza nota o nota a meno di un fattore di scala, tra due modelli che descrivono in maniera accettabile gli stessi dati vada meglio quello in cui l'indice $AIC = WRSS + 2M$ o $AIC = N \ln WRSS + 2M$ è minore.

Capitolo 4

Identificazione dei modelli

In questo capitolo viene riportata la procedura adottata per identificare i due modelli oggetto della nostra discussione: il modello “feedback-integrativo” proposto da Altafini e il modello “feed-forward” proposto in questa tesi. È stato usato lo stimatore ai minimi quadrati non lineari pesati e sono state elaborate delle strategie per cercare di avvicinarsi il più possibile al minimo globale della funzione costo e per cercare di migliorare l’affidabilità delle stime.

Nel corso del capitolo verranno riportate delle tabelle con i risultati della stima parametrica. Riprendendo alcuni concetti espressi nel paragrafo dedicato all’implementazione dei modelli (paragrafo 3.3), vogliamo qui spiegare il contenuto delle tabelle per favorire una migliore comprensione di ciò che verrà illustrato in questo capitolo. Ogni riga di ciascuna tabella contiene i risultati dell’identificazione parametrica ottenuta usando i dati di una serie temporale di un determinato feed-forward loop. Vengono raggruppate in cinque righe consecutive le serie temporali relative ad un singolo FFL. Nelle colonne sono riportati i valori delle stime ottenute e i vari indici che sono stati usati per la valutazione della bontà dei risultati e il confronto degli stessi.

In particolare per ogni parametro si riportano la stima iniziale dalla quale si è fatto partire l’algoritmo, la stima ottima ottenuta e il coefficiente di variazione (CV) come indice dell’incertezza associata alla stima. Per come è definito il CV, dei valori percentuali inferiori al 50 % indicano che il parametro è stimato con un buon grado di precisione.

Nelle tabelle compare poi il valore della deviazione standard dell’errore di misura calcolata a posteriori (indicata con l’abbreviazione sd_{post}). Si userà questo come indice di confronto della bontà di due identificazioni differenti fatte sugli stessi dati. La deviazione standard dell’errore è calcolata a posteriori come la radice del rapporto tra la funzione obiettivo calcolata nel punto di minimo e la differenza tra il numero di parametri e il numero di dati. Quindi un minor valore di SD è dovuto alla discesa della funzione obiettivo perciò indica il raggiungimento di un fit

migliore.

Il numero di corse, indicato con “nc”, rappresenta il numero di volte in cui i residui (distanza della predizione del modello dai dati) cambiano di segno ovvero rappresenta i passaggi dei residui per lo zero e quindi è una misura della cosiddetta “bianchezza” dei residui cioè di quanto siano posizionati in maniera random. Come spiegato nel paragrafo 3.3 i residui costituiscono una stima dell’errore di misura perciò è logico attendersi che siano compatibili con le proprietà statistiche dell’errore di misura. Essendo quest’ultimo a campioni scorrelati, ci aspettiamo dei residui sufficientemente random.

L’ultimo indice che compare nelle tabelle è il valore del coefficiente di Akaike (indicato con l’abbreviazione “AIC”). Verrà usato per confrontare modelli in competizione di ordine diverso ovvero il modello che meglio bilancia le esigenze contrastanti di fit e precisione sarà quello con Akaike minore.

Illustreremo dapprima il procedimento seguito per l’identificazione parametrica del modello “feed-forward” (paragrafi 4.1.1 e 4.1.2) e poi quello seguito per l’identificazione del modello “feedback-integrativo”(paragrafi 4.2.1 e 4.2.2).

4.1. Modello “feed-forward”

4.1.1 Stima di k_{31} e k_{32}

Consideriamo il modello che descrive il comportamento del feed-forward loop il cui pattern viene riportato in Figura 4.1:

$$\begin{cases} \dot{x}_2 = k_{21}x_1 - k_{02}x_2 \\ \dot{x}_3 = k_{31}x_1 + k_{32}x_2 - k_{03}x_3 \end{cases} \quad \begin{cases} x_2(0) = 0 \\ x_3(0) = 0 \end{cases}$$

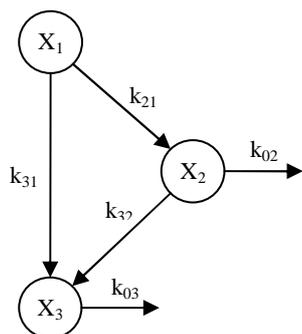


Figura 4.1 Grafo rappresentante un motivo di regolazione feed-forward.

I dati relativi ai compartimenti X_1 e X_2 sono stati considerati come funzioni forzanti, in questo modo il sistema si riduce alla sola seconda equazione differenziale: si è quindi voluto predire l'andamento del profilo di espressione del gene X_3 considerando i due regolatori, X_1 e X_2 (vedi Figura 4.2):

$$\dot{x}_3 = k_{31}x_1 + k_{32}x_2 - k_{03}x_3 \quad x_3(0) = 0 \quad (1)$$

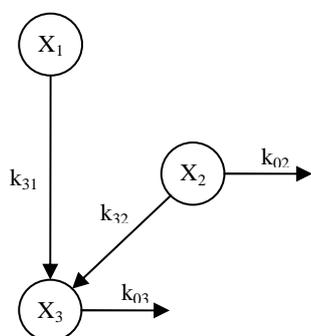


Figura 4.2 Grafo rappresentante il modello ridotto.

In questo modello, in un primo momento fissiamo il parametro di degradazione k_{03} al valore 0,019 che corrisponde al valor medio ottenuto nel lavoro di Altafini [3] fittando il modello con i dati di espressione e i dati proteici.

Procediamo all'identificazione del modello stimando quindi solamente i parametri k_{31} e k_{32} cioè i parametri che rappresentano la forza della regolazione.

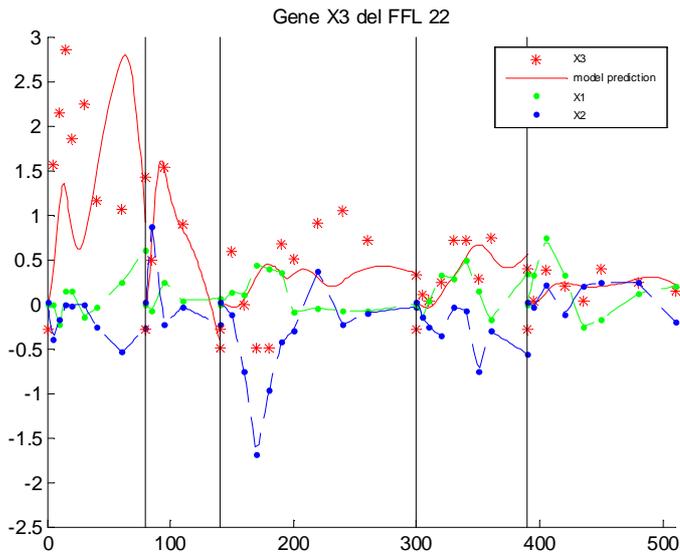
Consideriamo quindi la struttura del modello data dall'equazione differenziale (1) e utilizziamo, in un primo momento, i dati di espressione relativi ai FFL 22 e 25, di cui si è parlato nel Capitolo 3.

Il modello viene identificato per ogni serie temporale partendo da un paio di set di stime iniziali, scelte in modo casuale, per i due parametri da stimare. Questo primo step ci ha permesso di individuare il range di valori attorno ai quali si collocano i suddetti parametri, ovvero $[-2, 2]$. Abbiamo quindi ristretto i bound della funzione `lsqnonlin` al range appena trovato e abbiamo ripetuto la stima partendo da circa 60 set di stime iniziali diverse, scegliendo valori interni al range. Questo ci ha permesso di individuare quali stime iniziali conducono al fit migliore (non in assoluto bensì relativamente ai tentativi fatti).

Il fit migliore, ovvero la migliore aderenza del modello ai dati, lo valutiamo dalla norma al quadrato del vettore dei residui normalizzati, che corrisponde al valore della funzione obiettivo calcolata per il vettore ottimo dei parametri. Si sceglie dunque il modello con minore indice *WRSS* (si veda l'equazione numero 8 del capitolo 3). Per comodità, anziché valutare *WRSS*, abbiamo considerato la deviazione standard a posteriore che differisce solamente per un fattore di scala infatti:

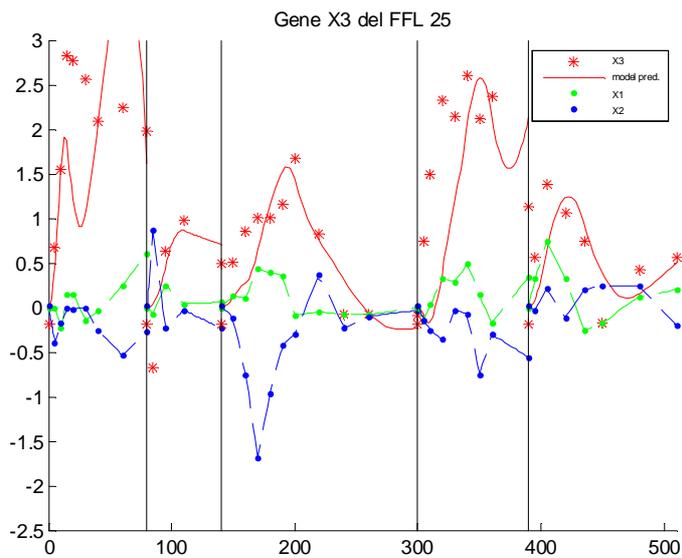
$$sd_{post} = \frac{WRSS(\hat{p})}{N - M}.$$

Nella Figura 4.3 vengono riportati i plot del fit migliore e le tabelle con i risultati in cui sono mostrate le stime ottime, i coefficienti di variazione associati ai parametri, il valore della deviazione standard a posteriori, il numero di corse e il valore del coefficiente di Akaike. Questi indici verranno sempre riportati in tutte le tabelle che mostreremo nel capitolo e in quello successivo.



FFL22

S	k_{31_iniz}	k_{31_ott}	CV_ k_{31}	k_{32_iniz}	k_{32_ott}	CV_ k_{32}	sd_post	nc	AIC
1	-2,	-0,4809	58	-0,4	-0,3873	35	1,2819	3	26
2	0,8	0,1315	42	0,9	0,3236	13	0,2009	2	-7
3	0,7	-0,0885	51	0,4	-0,0439	7	0,6002	4	17
4	-0,3	0,0749	4	0,08	-0,0066	8	0,275	4	-2
5	1,	0,0181	63	1,	0,0331	64	0,1896	4	-8



FFL25

S	k_{31_iniz}	k_{31_ott}	$CV_{k_{31}}$	k_{32_iniz}	k_{32_ott}	$CV_{k_{32}}$	sd_post	nc	AIC
1	-1,	-0,6388	24	-1,	-0,5452	12	1,1122	3	23
2	1,	0,3323	22	1,	0,0465	7	0,4602	2	2
3	0,4	0,1901	16	0,3	0,007	66	0,3413	5	5
4	0,5	0,2942	29	-0,02	-0,0227	13	1,1692	4	24
5	0,4	0,1134	23	0,15	-0,0038	76	0,4581	3	6

Figura 4.3 Plot dei dati e della predizione del modello (linea rossa continua) dei FFL 22 e 25. Sotto a ciascun plot sono riportate le tabelle con le stime iniziale e ottime dei parametri, i coefficienti di variazione (CV), la standard deviation a posteriori (sd_post), il numero di corse (nc) e il coefficiente di Akaike (AIC).

A questo punto abbiamo deciso di identificare il modello anche per altri feed-forward loop della rete del *S. cerevisiae*, come è stato discusso nel capitolo 3.

Anche in questo caso, per le 10 triplete scelte è stata fatta un'identificazione per ogni serie di ogni FFL.

Come abbiamo già visto, è necessario fornire all'algoritmo di ottimizzazione una stima iniziale. Si è scelto allora di costruire 4 set differenti di valori iniziali partendo dai valori ottimi ottenuti dall'identificazione appena descritta sui feed-forward loop 22 e 25.

I 4 set di stime iniziali sono riportati nella Tabella 1.

Nel primo set il valore per k_{31} è il risultato della media delle dieci stime ottime ottenute per k_{31} in tutte le serie temporali di entrambi i feed-forward loop. Analogamente è stata calcolata la media dei dieci valori relativi a k_{32} .

Nel secondo set sono state prese semplicemente le stime ottime ottenute considerando il feed-forward loop 22.

La stessa cosa è stata fatta per il terzo set considerando però il feed-forward loop 25. Infine nel quarto set è stata considerata la media, calcolata per ogni serie, dei due valori ottimi ottenuti per il parametro k_{31} sui due FFL. Analogamente si è ricavata la media per i due valori relativi a k_{32} .

SET	SERIE	k_{31}	k_{32}
1	1,2,3,4,5	-0,0054	0,0599
2	1	-0,4809	-0,3873
	2	0,1315	0,3236
	3	-0,0885	-0,0439
	4	0,0749	-0,0066
	5	0,0181	0,0331
3	1	-0,6388	-0,5452
	2	0,3323	0,0465
	3	0,1901	0,007
	4	0,2942	-0,0227
	5	0,1134	-0,0038
4	1	-0,5599	-0,4663
	2	0,2319	0,1851
	3	0,0508	-0,0184
	4	0,1845	-0,0146
	5	0,0658	0,0146

Tabella 14 set di stime iniziali per i parametri k_{31} k_{32} .

Ogni set di stime iniziali è stato dunque fornito all'algoritmo per l'identificazione del modello usando i dati dei 10 feed-forward loop.

Anche qui sono stati calcolati, per ogni identificazione, la deviazione standard a posteriori e il coefficiente di Akaike. A questo punto è stata scelta, per ognuno dei 50 casi (5 serie per 10 FFL), l'identificazione migliore (WRSS minore) tra le quattro ottenute partendo dai quattro set di valori iniziali (Tabella 2).

FFL	Serie	k_{31_iniz}	k_{31_ott}	CV_ k_{31}	k_{32_iniz}	k_{32_ott}	CV_ k_{32}	sd_post	nc	AIC
4	1	-0,0054	-0,2336	19	-0,0599	0,3569	18	0,5448	7	11
	2	-0,0054	0,1217	80	-0,0599	0,1127	86	0,5156	3	3
	3	-0,0054	-0,028	68	-0,0599	-0,014	190	0,8006	9	23
	4	-0,0054	-0,0942	46	-0,0599	0,038	25	0,2922	5	-1
	5	-0,0054	-0,0037	558	-0,0599	0,1941	17	0,2588	3	-3

5	1	-0,0054	-0,1977	15	-0,0599	0,2749	17	0,4153	7	6
	2	-0,0054	0,0734	13	-0,0599	0,0441	20	0,0764	2	-16
	3	0,0508	-0,0618	11	-0,0184	0,0223	50	0,2528	7	-2
	4	-0,0054	-0,1149	43	-0,0599	0,0439	39	0,434	5	6
	5	0,0658	-0,0213	99	0,0146	0,199	16	0,2525	6	-4
7	1	-0,0054	-0,0076	684	-0,0599	-0,0211	294	0,3429	8	2
	2	-0,0054	0,0625	129	-0,0599	-0,046	93	0,4808	4	2
	3	-0,0054	0,1867	0	-0,0599	-0,0259	0	0,372	7	6
	4	-0,0054	-0,0476	6	-0,0599	-0,0284	10	0,1079	5	-19
	5	-0,0054	0,0008	1.851	-0,0599	0,0379	64	0,152	6	-12
9	1	-0,0054	-0,2071	12	-0,0599	0,1941	23	0,5169	6	10
	2	-0,0054	0,2642	30	-0,0599	0,1409	62	0,5091	4	3
	3	0,1901	0,1814	25	0,007	0,0672	76	0,6706	5	19
	4	-0,0054	-0,0143	717	-0,0599	-0,5628	18	0,823	5	18
	5	0,0658	0,0218	86	0,0146	-0,0974	61	0,7345	6	13
11	1	-0,0054	-0,356	0	-0,0599	-0,3203	11	0,6825	3	15
	2	-0,0054	0,2569	33	-0,0599	0,0487	122	0,3034	4	-2
	3	0,0508	-0,1187	0	-0,0184	-0,0301	0	0,5486	6	15
	4	-0,0054	0,1262	28	-0,0599	0,006	51	0,2088	4	-7
	5	-0,0054	0,0101	158	-0,0599	0,069	0	0,2236	6	-6
29	1	-0,0054	-0,0024	1.107	-0,0599	-0,0159	30	0,1901	5	-8
	2	-0,0054	-0,0708	1.199	-0,0599	-0,0044	1722	0,511	3	3
	3	-0,0054	-0,0095	374	-0,0599	-0,0638	25	0,5765	8	16
	4	-0,0054	-0,0053	666	-0,0599	-0,0166	87	0,2447	7	-4
	5	-0,0054	-0,0066	413	-0,0599	0,0159	169	0,2015	5	-7
31	1	-0,0054	-0,1429	13	-0,0599	0,0541	2	0,6352	4	13
	2	-0,0054	0,1635	68	-0,0599	-0,078	36	0,3491	4	-1
	3	0,0508	0,076	81	-0,0184	-0,0161	17	0,7251	5	21
	4	-0,0054	0,5767	23	-0,0599	-0,1055	17	0,6082	4	13
	5	-0,0054	-0,0067	674	-0,0599	0,0002	2555	0,2283	6	-5
32	1	-0,0054	0,0359	193	-0,0599	-0,0232	146	0,6198	7	13
	2	-0,0054	-0,0081	556	-0,0599	-0,0095	472	0,5294	3	3
	3	-0,0054	0,0255	46	-0,0599	-0,0352	10	0,7	8	20
	4	-0,0054	-0,207	8	-0,0599	0,0726	22	0,398	7	5
	5	-0,0054	0,0184	581	-0,0599	-0,1159	106	0,4999	6	7
41	1	-0,0054	0,386	21	-0,0599	-0,0221	197	0,7007	5	15
	2	-0,0054	-0,0529	572	-0,0599	-0,0275	1571	0,8374	2	8
	3	-0,0054	0,0134	42	-0,0599	-0,0098	41	0,6404	8	18
	4	-0,0054	-0,0108	34	-0,0599	0,0568	46	0,2715	5	-2
	5	-0,0054	0,0019	745	-0,0599	0,0435	51	0,095	6	-19
50	1	-0,0054	-0,0525	58	-0,0599	0,0643	48	0,3676	7	4
	2	-0,0054	-0,3204	59	-0,0599	0,0632	56	0,3177	4	-2
	3	-0,0054	0,0219	96	-0,0599	0,0018	669	0,1717	5	-11
	4	-0,0054	-0,1054	30	-0,0599	0,0804	19	0,3699	5	4
	5	-0,0054	0,0047	237	-0,0599	0,0094	63	0,1704	5	-10

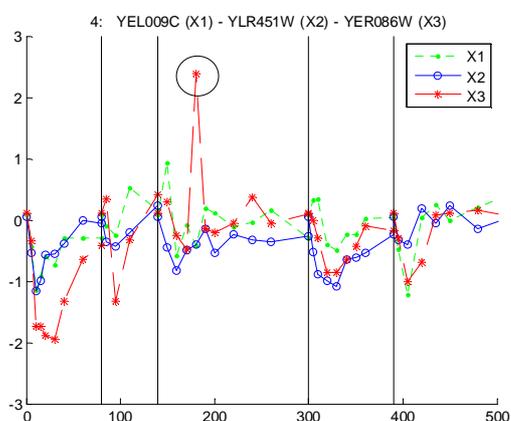
Tabella 2

Per ridurre ulteriormente la probabilità di essersi imbattuti in minimi locali della funzione costo, si è deciso di considerare altre stime iniziali da cui far partire la procedura di ottimizzazione (lsqnonlin) che sono state scelte secondo il procedimento che verrà ora illustrato.

Consideriamo, per ogni coppia di stime ottime (k_{31_ott} e k_{32_ott}), due vettori: il primo vettore è costituito da valori numerici presi in maniera casuale in un intorno di k_{31_ott} , il secondo vettore è costruito in modo analogo considerando un intorno di k_{32_ott} . Vengono poi individuate tutte le possibili coppie di valori e si riparte a fare la stima dei parametri da ognuna di queste coppie di stime iniziali.

A questo punto del lavoro ci si è resi conto che alcune volte il fit risultava particolarmente “cattivo” molto probabilmente a causa del fatto che uno dei campioni era evidentemente molto rumoroso. Questa considerazione è nata da un’attenta ispezione visiva del plot dei dati a disposizione, durante la quale si è notato che il campione in questione era nettamente distante dai campioni adiacenti. Se si osserva infatti il trend disegnato dall’interpolazione dei dati di altre serie si possono facilmente riconoscere dei profili temporali piuttosto coerenti, cosa che invece non appare nei seguenti casi: nella serie numero 3 del feed-forward loop 4, nella serie numero 3 del feed-forward loop 31 e nella serie numero 1 del feed-forward loop 41.

Nei primi due casi il campione molto rumoroso è il quinto campione relativo al gene regolato X_3 . Nel terzo caso si tratta del settimo campione relativo al gene regolatore X_2 (Figura 4.4).



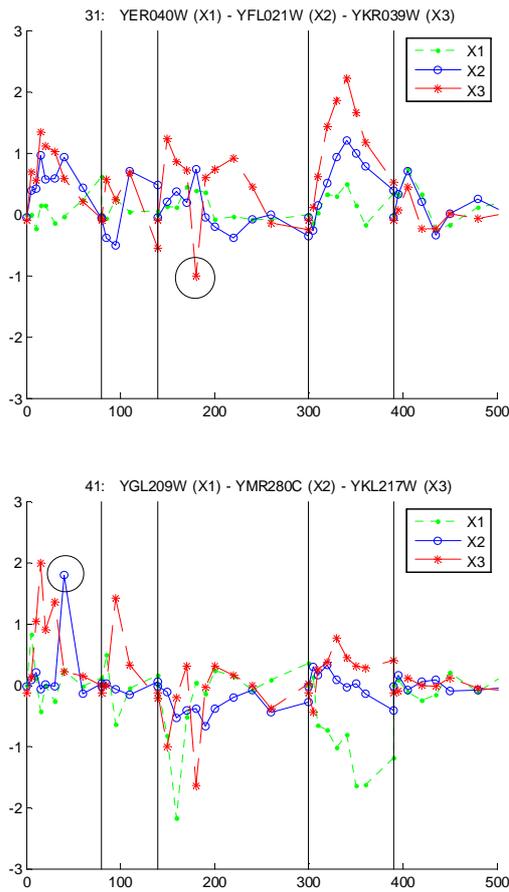


Figura 4.4

Si è deciso quindi di ripetere l'identificazione togliendo i suddetti campioni. I risultati migliorano perciò ora considereremo sempre (anche durante l'identificazione del modello Altafini nei paragrafi successivi) i dati privati di questi 3 campioni.

Nella Tabella 3 vengono riportati i risultati ottenuti dopo la procedura appena illustrata e dopo aver tolto i 3 campioni rumorosi.

FFL	Serie	k_{31_iniz}	k_{31_ott}	$CV_{k_{31}}$	k_{32_iniz}	k_{32_ott}	$CV_{k_{32}}$	sd_post	nc	AIC
4	1	-1,	-0,2153	18	1,	0,3382	18	0,5437	7	11
	2	-0,0054	0,1217	80	-0,0599	0,1127	86	0,5156	3	3
	3	-0,8	-0,0551	27	-0,4	-0,0131	177	0,7968	9	23
	4	-0,0054	-0,0942	46	-0,0599	0,038	25	0,2922	5	-1
	5	-1,	-0,0043	576	0,5	0,1967	25	0,2585	3	-3
5	1	-0,0054	-0,1977	15	-0,0599	0,2749	17	0,4153	7	6
	2	-0,0054	0,0734	13	-0,0599	0,0441	20	0,0764	2	-16
	3	0,0508	-0,0618	11	-0,0184	0,0223	50	0,2528	7	-2
	4	-0,0054	-0,1149	43	-0,0599	0,0439	39	0,434	5	6
	5	0,0658	-0,0213	99	0,0146	0,199	16	0,2525	6	-4

7	1	0,01	0,0044	1.321	-0,05	-0,0334	192	0,3425	8	2
	2	-0,0054	0,0625	129	-0,0599	-0,046	93	0,4808	4	2
	3	0,5	0,1862	2	-0,1	-0,0255	12	0,3717	7	6
	4	-0,0054	-0,0476	6	-0,0599	-0,0284	10	0,1079	5	-19
	5	0,001	0,0041	360	0,04	0,0237	165	0,1507	6	-12
9	1	-0,0054	-0,2071	12	-0,0599	0,1941	23	0,5169	6	10
	2	-0,0054	0,2642	30	-0,0599	0,1409	62	0,5091	4	3
	3	0,8	0,1909	7	0,1	0,1379	24	0,6118	6	17
	4	0,1	0,0075	1.209	-0,2	-0,5746	18	0,8166	5	18
	5	0,04	0,0305	59	0,2	-0,078	60	0,7329	6	13
11	1	-1,	-0,3218	44	-1,	-0,2858	21	0,6799	3	15
	2	-0,0054	0,2569	33	-0,0599	0,0487	122	0,3034	4	-2
	3	-0,5	-0,1857	58	-0,8	-0,0544	39	0,5418	8	15
	4	-0,0054	0,1262	28	-0,0599	0,006	51	0,2088	4	-7
	5	-0,0054	0,0101	158	-0,0599	0,069	0	0,2236	6	-6
29	1	-0,5	-0,0025	1.068	-0,02	-0,0157	26	0,19	5	-8
	2	-0,09	-0,0706	317	0,1	-0,0045	306	0,511	3	3
	3	-0,1	-0,0285	101	-0,02	-0,055	2	0,5733	8	16
	4	-0,0054	-0,0053	666	-0,0599	-0,0166	87	0,2447	7	-4
	5	-0,5	-0,0054	720	0,5	0,0146	304	0,2013	5	-7
31	1	-0,0054	-0,1429	13	-0,0599	0,0541	2	0,6352	4	13
	2	-0,0054	0,1635	68	-0,0599	-0,078	36	0,3491	4	-1
	3	0,0508	0,076	81	-0,0184	-0,0161	17	0,7251	5	21
	4	-0,0054	0,5767	23	-0,0599	-0,1055	17	0,6082	4	13
	5	-0,1	-0,0225	198	0,01	0,017	281	0,2266	6	-5
32	1	-0,0054	0,0359	193	-0,0599	-0,0232	146	0,6198	7	13
	2	-0,0054	-0,0081	556	-0,0599	-0,0095	472	0,5294	3	3
	3	-0,0054	0,0255	46	-0,0599	-0,0352	10	0,7	8	20
	4	-0,0054	-0,207	8	-0,0599	0,0726	22	0,398	7	5
	5	-0,0054	0,0184	581	-0,0599	-0,1159	106	0,4999	6	7
41	1	-0,0054	0,386	21	-0,0599	-0,0221	197	0,7007	5	15
	2	-1,	-0,0738	397	-0,5	0,0391	672	0,8351	2	8
	3	0,1	0,0172	108	-0,1	-0,0026	496	0,631	6	18
	4	-0,0054	-0,0108	34	-0,0599	0,0568	46	0,2715	5	-2
	5	-0,5	0,0023	584	0,04	0,0464	23	0,0949	5	-19
50	1	-0,0054	-0,0525	58	-0,0599	0,0643	48	0,3676	7	4
	2	-0,0054	-0,3204	59	-0,0599	0,0632	56	0,3177	4	-2
	3	-0,0054	0,0219	96	-0,0599	0,0018	669	0,1717	5	-11
	4	-0,0054	-0,1054	30	-0,0599	0,0804	19	0,3699	5	4
	5	-0,0054	0,0047	237	-0,0599	0,0094	63	0,1704	5	-10

Tabella 3

Da questo punto in poi illustriamo tutto l'iter che è stata seguito allo scopo di cercare di migliorare alcuni dei risultati che, come verrà spiegato, sono ritenuti poco soddisfacenti. Nella Tabella 3 ci sono dei casi in cui il coefficiente di variazione, che fornisce una misura della precisione con cui stimiamo i parametri, assume valori molto alti e ciò significa che stiamo

stimando con un'incertezza troppo elevata. È stato scelto un valore massimo pari a 150 % oltre al quale le precisioni delle stime si considerano troppo incerte.

A questo punto si è deciso di fare nuovi tentativi di stima, spostando i valori iniziali, solo per i casi individuati, per cercare di ottenere dei risultati più soddisfacenti in termini di CV. A questo scopo è stata calcolata la media delle stime ottime, considerando solo i casi in cui le precisioni delle stime sono ritenute sufficienti (<150) o valide (cioè diverse da zero).

I valori medi ottenuti sono:

$$k_{31} = 0.0039$$

$$k_{32} = 0.0436.$$

Si è allora fatta ripartire l'identificazione dando come stime iniziali tali valori medi. In alcuni casi i CV migliorano ma, a volte, a discapito del valore della deviazione standard a posteriori, ovvero il fit peggiora. Aggiorniamo quindi la nostra tabella inserendo solo in casi in cui il fit è migliorato o è rimasto inalterato e i CV sono migliorati.

Ci sono ancora dei CV elevati. Allora ricalcoliamo le medie, escludendo sempre i casi con CV alti o pari a zero. I nuovi valori medi sono:

$$k_{31} = 0.0013$$

$$k_{32} = 0.062$$

Solo per i casi da migliorare, fissiamo il parametro k_{31} al valore medio e stimiamo solo il parametro k_{32} facendolo partire da 0.062.

Facciamo anche il contrario, ovvero fissiamo k_{32} e stimiamo solo k_{31} facendolo partire da 0.0013. Otteniamo qualche miglioramento (Tabella 4).

FFL	Serie	k_{31_iniz}	k_{31_ott}	CV_ k_{31}	k_{32_iniz}	k_{32_ott}	CV_ k_{32}	sd_post	nc	AIC
4	1	-1,	-0,2298	20	1,	0,3508	19	0,5497	3	10,74
	2	-0,0054	0,1223	79	-0,0599	0,1117	87	0,517	3	2,90
	3		0,0013		0,062	0,0036	2	0,2541	4	-3,43
	4	-0,0054	-0,1087	24	-0,0599	0,0398	15	0,2936	2	-0,55

	5	0,0039	0,0003	47	0,0436	0,1698	0	0,2679	2	-2,74
5	1	-0,0054	-0,1252	21	-0,0599	0,1995	21	0,4619	2	7,61
	2	-0,0054	0,0722	12	-0,0599	0,0428	19	0,0754	2	-16,35
	3	0,0508	-0,061	12	-0,0184	0,0219	30	0,2533	6	-2,04
	4	-0,0054	-0,016	60	-0,0599	0,0247	33	0,4988	2	8,99
	5	0,0658	-0,0261	66	0,0146	0,2154	9	0,2532	4	-3,64
7	1		0,0013		0,062	-0,0292	30	0,3204	4	0,23
	2	-0,0054	0,0616	128	-0,0599	-0,0447	94	0,4813	2	2,18
	3	0,5	0,2016	2	-0,1	-0,0262	13	0,3735	7	6,51
	4	-0,0054	-0,0477	29	-0,0599	-0,0274	17	0,105	3	-19,05
	5		0,0013		0,062	0,0315	20	0,1399	6	-13,90
9	1	-0,0054	-0,2016	6	-0,0599	0,1886	20	0,5236	2	9,87
	2	-0,0054	0,2613	30	-0,0599	0,1382	63	0,5134	2	2,83
	3	0,8	0,1937	5	0,1	0,1538	15	0,6126	2	17,39
	4		0,0013		0,062	-0,5599	12	0,7432	3	15,37
	5	0,04	0,0278	68	0,2	-0,0802	64	0,7384	5	13,48
11	1	-1,	-0,3257	44	-1,	-0,2867	22	0,6786	3	14,54
	2	-0,0054	0,2523	33	-0,0599	0,0482	120	0,3015	3	-2,50
	3	-0,5	-0,1703	0	-0,8	-0,0504	0	0,5404	4	14,63
	4	0,0039	0,1337	21	0,0436	0,0136	145	0,2115	4	-6,45
	5	0,0039	0,009	0	0,0436	0,069	5	0,2269	6	-5,40
29	1		0,0013		0,062	-0,0161	20	0,1774	4	-10,41
	2		0,0013		0,062	-0,0147	221	0,4556	1	1,07
	3	-0,1	-0,0257	135	-0,02	-0,0561	11	0,5748	3	15,99
	4	-0,0066	-0,0142	125	0,0387	-0,0135	101	0,2448	5	-3,82
	5		0,0013		0,062	0,0082	175	0,1866	5	-9,29
31	1	-0,0054	-0,1439	63	-0,0599	0,0496	32	0,634	3	13,31
	2	-0,0054	0,1606	68	-0,0599	-0,0778	36	0,3483	4	-1,06
	3	0,0508	0,1063	38	-0,0184	0,0034	648	0,5396	4	12,46
	4	-0,0054	0,5621	23	-0,0599	-0,1023	14	0,6105	4	12,63
	5		0,0013		0,062	-0,0073	187	0,2132	5	-7,16
32	1		0,0013		0,062	-0,0191	87	0,5859	4	11,09
	2		0,0013		0,062	-0,0106	316	0,4585	3	1,13
	3		0,0013		0,062	-0,0221	5	0,6711	3	18,56
	4	0,0039	-0,208	11	0,0436	0,0727	32	0,4013	3	5,08
	5		0,0013		0,062	-0,0963	32	0,4662	5	5,36
41	1	0,0039	0,0125	82	0,0436	0,4709	61	0,8899	1	19,41
	2	0,0013	-0,0839	27		0,062		0,7242	2	5,70
	3		0,0013		0,062	0,0092	197	0,6267	5	17,05
	4	-0,0054	-0,0105	35	-0,0599	0,0573	45	0,2718	3	-1,94
	5		0,0013		0,062	0,0441	54	0,0878	3	-21,35
50	1	-0,0054	-0,0528	58	-0,0599	0,0645	47	0,3679	3	3,52
	2	-0,0054	-0,3215	59	-0,0599	0,0628	57	0,3184	2	-1,95
	3		0,0013		0,062	0,0133	11	0,1671	3	-12,04
	4	-0,0054	-0,1049	30	-0,0599	0,0794	19	0,3728	3	3,75
	5		0,0013		0,062	0,0099	69	0,1597	3	-11,78

Tabella 4

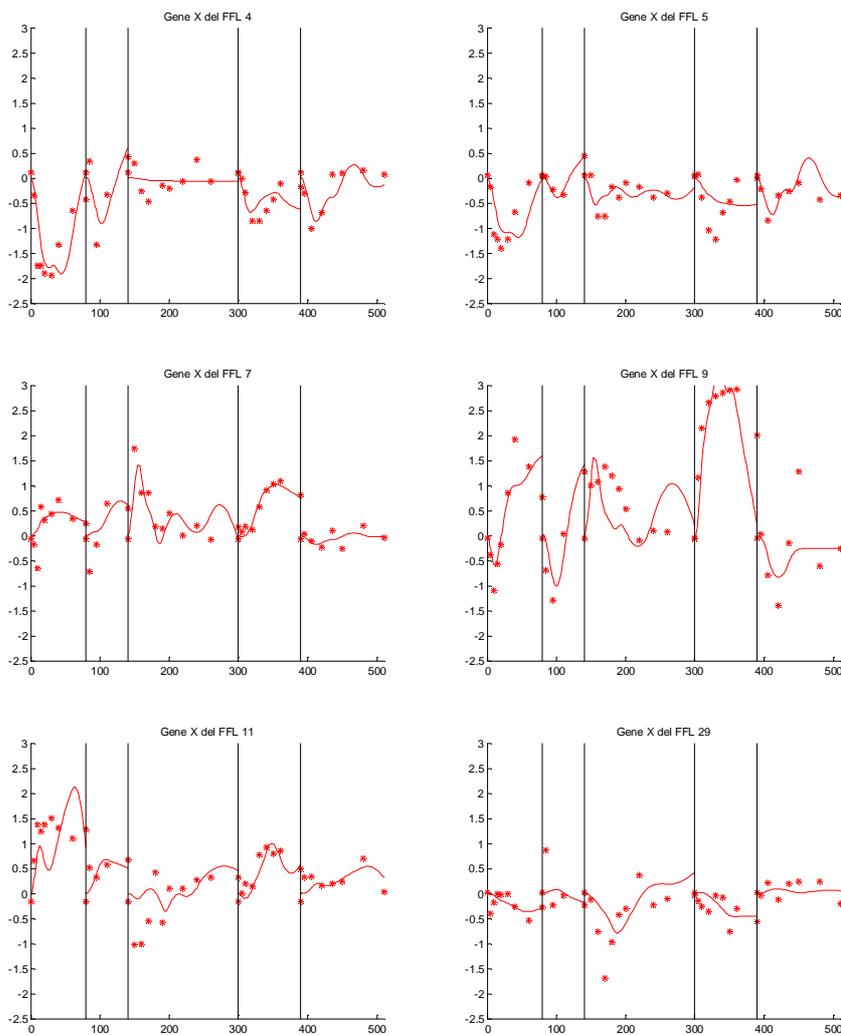
Infine, per gli ultimi casi rimasti con $CV > 150\%$, proviamo a fare alcuni tentativi di stima, solo per questi casi, spostando i valori iniziali di qualche centesimo. Così facendo otteniamo altri miglioramenti. Di nuovo rifacciamo le medie escludendo i casi ancora rimasti con CV elevati o zero, e ripetiamo l'identificazione. Non otteniamo miglioramenti. Fissiamo allora un parametro e poi l'altro. Alla fine di tutti i passaggi si ottengono i risultati riportati nella Tabella 5.

FFL	Serie	k_{31_iniz}	k_{31_ott}	$CV_{k_{31}}$	k_{32_iniz}	k_{32_ott}	$CV_{k_{32}}$	sd_post	nc	AIC
4	1	-1,	-0,2298	20	1,	0,3508	19	0,5497	3	10,74
	2	-0,0054	0,1223	79	-0,0599	0,1117	87	0,517	3	2,90
	3		0,0013		0,062	0,0036	2	0,2541	4	-3,43
	4	-0,0054	-0,1087	24	-0,0599	0,0398	15	0,2936	2	-0,55
	5	0,0039	0,0003	47	0,0436	0,1698	0	0,2679	2	-2,74
5	1	-0,0054	-0,1252	21	-0,0599	0,1995	21	0,4619	2	7,61
	2	-0,0054	0,0722	12	-0,0599	0,0428	19	0,0754	2	-16,35
	3	0,0508	-0,061	12	-0,0184	0,0219	30	0,2533	6	-2,04
	4	-0,0054	-0,016	60	-0,0599	0,0247	33	0,4988	2	8,99
	5	0,0658	-0,0261	66	0,0146	0,2154	9	0,2532	4	-3,64
7	1		0,0013		0,062	-0,0292	30	0,3204	4	0,23
	2	-0,0054	0,0616	128	-0,0599	-0,0447	94	0,4813	2	2,18
	3	0,5	0,2016	2	-0,1	-0,0262	13	0,3735	7	6,51
	4	-0,0054	-0,0477	29	-0,0599	-0,0274	17	0,105	3	-19,05
	5		0,0013		0,062	0,0315	20	0,1399	6	-13,90
9	1	-0,0054	-0,2016	6	-0,0599	0,1886	20	0,5236	2	9,87
	2	-0,0054	0,2613	30	-0,0599	0,1382	63	0,5134	2	2,83
	3	0,8	0,1937	5	0,1	0,1538	15	0,6126	2	17,39
	4		0,0013		0,062	-0,5599	12	0,7432	3	15,37
	5	0,04	0,0278	68	0,2	-0,0802	64	0,7384	5	13,48
11	1	-1,	-0,3257	44	-1,	-0,2867	22	0,6786	3	14,54
	2	-0,0054	0,2523	33	-0,0599	0,0482	120	0,3015	3	-2,50
	3	-0,2	-0,1764	30	-0,05	-0,0513	16	0,5399	4	14,61
	4	0,0039	0,1337	21	0,0436	0,0136	145	0,2115	4	-6,45
	5	0,01	0,0105	79	0,05	0,0654	5	0,2265	6	-5,43
29	1		0,0013		0,062	-0,0161	20	0,1774	4	-10,41
	2		-0,0019		0,0277	-0,014	215	0,4544	1	1,04
	3	-0,1	-0,0257	135	-0,02	-0,0561	11	0,5748	3	15,99
	4	-0,0066	-0,0142	125	0,0387	-0,0135	101	0,2448	5	-3,82
	5		-0,0019		0,0277	0,012	58	0,1865	5	-9,30
31	1	-0,0054	-0,1439	63	-0,0599	0,0496	32	0,634	3	13,31
	2	-0,0054	0,1606	68	-0,0599	-0,0778	36	0,3483	4	-1,06
	3	0,2	0,1044	61	0,005	0,0061	53	0,5407	4	12,50
	4	-0,0054	0,5621	23	-0,0599	-0,1023	14	0,6105	4	12,63
	5		0,0013		-0,01	-0,0095	71	0,2132	5	-7,16
32	1		0,0013		0,062	-0,0191	87	0,5859	4	11,09
	2		0,0013		0,062	-0,0106	316	0,4585	3	1,13
	3		0,0013		0,062	-0,0221	5	0,6711	3	18,56
	4	0,0039	-0,208	11	0,0436	0,0727	32	0,4013	3	5,08
	5		0,0013		0,062	-0,0963	32	0,4662	5	5,36

41	1	0,0039	0,0125	82	0,0436	0,4709	61	0,8899	1	19,41
	2	0,0013	-0,0839	27		0,062		0,7242	2	5,70
	3		0,0013		0,005	0,0074	92	0,6269	5	17,05
	4	-0,0054	-0,0105	35	-0,0599	0,0573	45	0,2718	3	-1,94
	5		0,0013		0,062	0,0441	54	0,0878	3	-21,35
50	1	-0,0054	-0,0528	58	-0,0599	0,0645	47	0,3679	3	3,52
	2	-0,0054	-0,3215	59	-0,0599	0,0628	57	0,3184	2	-1,95
	3		0,0013		0,062	0,0133	11	0,1671	3	-12,04
	4	-0,0054	-0,1049	30	-0,0599	0,0794	19	0,3728	3	3,75
	5		0,0013		0,062	0,0099	69	0,1597	3	-11,78

Tabella 5

I plot del fit della predizione del modello contro i dati sono riportati nella Figura 4.5.



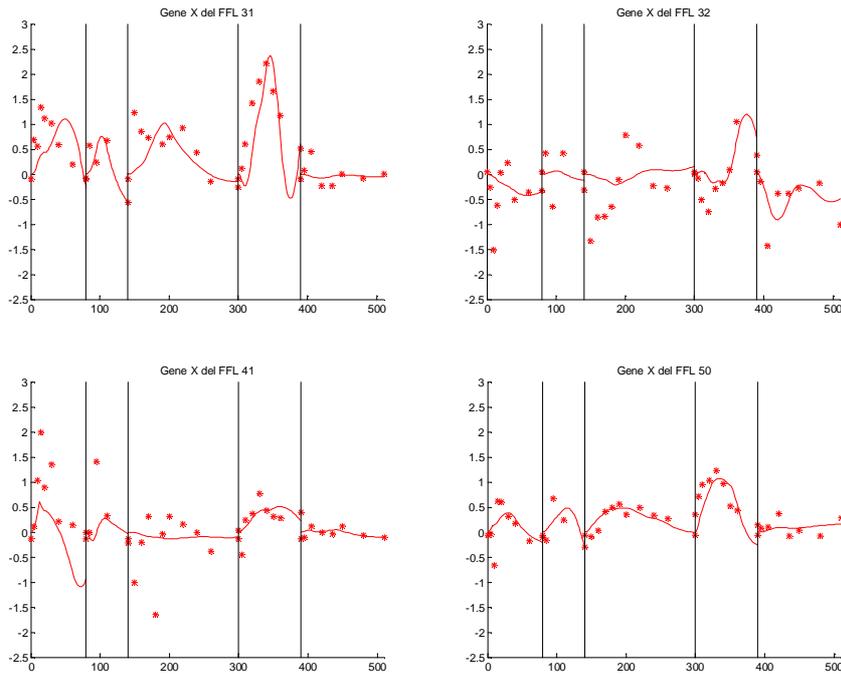
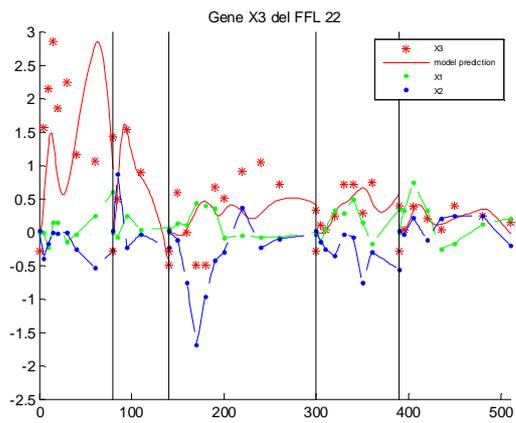


Figura 4.5

4.1.2. Stima di k_{31} , k_{32} e k_{03} .

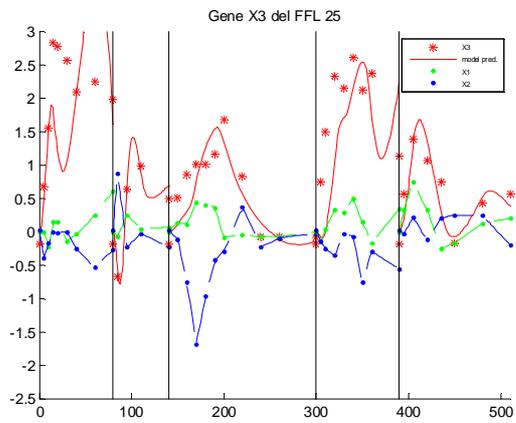
Ora vogliamo stimare anche il parametro di degradazione k_{03} che nel precedente paragrafo era stato fissato. Procediamo quindi con la stima dei tre parametri k_{31} , k_{32} e k_{03} .

In un primo momento eseguiamo l'identificazione usando i dati di espressione relativi ai FFL 22 e 25 partendo da circa 60 set di stime iniziali diverse. Tra queste abbiamo quindi scelto quelle che conducono al fit migliore sempre valutando la minore deviazione standard a posteriori. Nella Figura 4.6 sono riportati la figura con il plot dei dati verso la predizione del modello e la tabella con le stime dei parametri e i vari indici.



FFL22

S	k_{31_iniz}	k_{31_ott}	CV_k_{31}	k_{32_iniz}	k_{32_ott}	CV_k_{32}	k_{03_iniz}	k_{03_ott}	CV_k_{03}	sd_post	nc	AIC
1	-0,005	-0,6056	65	-0,06	-0,622	68	0,02	0,0471	76	1,3667	3	28
2	0,03	0,0244	12.357	-0,5	0,3077	150	0,01	0,0087	3748	0,2448	2	-5
3	-0,1	-0,1152	0	-0,06	-0,052	20	0,02	0,02	25	0,6383	4	19
4	0,5	0,1098	63	0,5	-0,0518	58	0,02	0,0727	53	0,2761	6	-1
5	-0,005	0,0601	50	-0,06	0,1634	71	0,02	0,1219	43	0,184	5	-8



FFL25

S	k_{31_iniz}	k_{31_ott}	CV_k_{31}	k_{32_iniz}	k_{32_ott}	CV_k_{32}	k_{03_iniz}	k_{03_ott}	CV_k_{03}	sd_post	nc	AIC
1	-0,5	-0,633	37	-0,5	-0,5441	44	0,02	0,0205	216	1,201	3	25
2	0,5	1,4394	4	-0,5	-0,234	38	0,05	0,1866	24	0,2127	2	-6
3	-0,5	0,1704	50	0,5	-0,0097	216	0,02	0,0306	44	0,3553	5	6
4	0,5	0,4505	71	-0,5	-0,2051	38	0,05	0,0773	73	1,0552	4	23
5	0,5	0,3107	27	-0,5	0,1855	49	0,05	0,1084	34	0,3158	5	0

Figura 4.6 Plot dei dati e della predizione del modello (linea rossa continua) dei FFL 22 e 25. Sotto a ciascun plot sono riportate le tabelle con le stime iniziale e ottime dei parametri, i coefficienti di variazione (CV), la standard deviation a posteriori (sd_post), il numero di corse (nc) e il coefficiente di Akaike (AIC).

A questo punto abbiamo deciso di identificare il modello anche per gli altri 10 feed-forward loop, come abbiamo fatto nel precedente paragrafo. Prima di tutto sono stati tolti i 3 campioni particolarmente rumorosi, poi sono stati costruiti 4 set differenti di stime iniziali partendo dai valori ottimi ottenuti dall'identificazione appena descritta sui feed-forward loop 22 e 25.

I 4 set di stime iniziali sono riportati nella Tabella 6. Per la costruzione del set si rimanda al paragrafo precedente.

SET	SERIE	k_{31}	k_{32}	k_{03}
1	1,2,3,4,5	0,1187	-0,1072	0,0695
2	1	-0,6056	-0,622	0,0471
	2	0,0244	0,3077	0,0087
	3	-0,1152	-0,052	0,02
	4	0,1098	-0,0518	0,0727
	5	0,0601	0,1634	0,1219
3	1	-0,633	-0,5441	0,0205
	2	1,4394	-0,234	0,1866
	3	0,1458	-0,0197	0,0322
	4	0,4505	-0,2051	0,0773
	5	0,3107	0,1855	0,1084
4	1	-0,6193	-0,5831	0,0338
	2	0,7319	0,0369	0,0976
	3	0,0153	-0,0359	0,0261
	4	0,2801	-0,1284	0,075
	5	0,1854	0,1745	0,1151

Tabella 6. 4 set di stime iniziali per i parametri k_{31} , k_{32} e k_{03} .

Ogni set di stime iniziali è stato dunque fornito all'algoritmo per l'identificazione del modello. A questo punto è stata scelta, per ognuno dei 50 casi, l'identificazione migliore tra le quattro ottenute partendo dai quattro set di valori iniziali (Tabella 7).

FFL	S	k _{31_iniz}	k _{31_ott}	CV_k ₃₁	k _{32_iniz}	k _{32_ott}	CV_k ₃₂	k _{03_iniz}	k _{03_ott}	CV_k ₀₃	sd_post	nc	AIC
4	1	0,1187	0,2504	81	-0,1072	0,1543	90	0,0695	0,1566	37	0,2376	4	-3,75
	2	0,0244	0,2092	101	0,3077	0,3556	83	0,0087	0,1515	120	0,4948	3	2,43
	3	0,0153	-0,0227	265	-0,0359	0,023	542	0,0261	0,0933	602	0,2801	2	0,00
	4	0,4505	0,14	26	-0,2051	0,1978	16	0,0773	0,3163	2	0,1489	5	-12,16
	5	0,1187	0,0351	73	-0,1072	0,2448	14	0,0695	0,0624	22	0,1362	4	-13,03
5	1	-0,6056	-0,0741	86	-0,622	0,3375	25	0,0471	0,1532	23	0,1611	5	-10,73
	2	0,1187	0,0704	14	-0,1072	0,0323	67	0,0695	0,	5,E+10	0,0702	2	-17,10
	3	0,1187	0,0917	3	-0,1072	0,1513	0	0,0695	0,2026	4	0,2288	6	-3,57
	4	0,2801	0,2781	23	-0,1284	0,3274	60	0,075	0,4255	54	0,2213	2	-5,02
	5	0,1187	-0,0158	285	-0,1072	0,267	36	0,0695	0,045	48	0,2804	3	-1,47
7	1	0,1187	0,019	443	-0,1072	-0,045	222	0,0695	0,0123	362	0,3707	4	4,26
	2	0,0244	0,4761	1.163	0,3077	0,0244	3710	0,0087	0,2317	877	0,5055	1	2,64
	3	0,1458	0,185	26	-0,0197	-0,0213	52	0,0322	0,0171	33	0,393	7	8,33
	4	0,1187	-0,0442	5	-0,1072	-0,0249	36	0,0695	0,015	38	0,107	4	-18,11
	5	0,1187	0,0762	18	-0,1072	-0,1316	24	0,0695	0,2091	47	0,1281	6	-14,00
9	1	0,1187	-0,3639	8	-0,1072	0,3502	15	0,0695	0,0592	30	0,416	6	6,34
	2	1,4394	0,5921	30	-0,234	0,4722	44	0,1866	0,1152	40	0,251	3	-4,36
	3	-0,1152	-0,0322	74	-0,052	0,1569	67	0,02	0,0479	132	0,8235	2	24,60
	4	0,1098	0,1804	13	-0,0518	-0,4833	10	0,0727	0,001	738	0,3815	3	4,78
	5	0,1854	0,0523	0	0,1745	-0,3691	10	0,1151	0,0943	29	0,5151	5	8,26
11	1	0,1187	-0,307	61	-0,1072	-0,4453	47	0,0695	0,0561	25	0,7533	3	17,03
	2	0,1187	0,1413	227	-0,1072	0,0637	109	0,0695	6,E-11	8,E+10	0,3236	3	-1,82
	3	-0,1152	-0,1853	5	-0,052	-0,0444	0	0,02	0,0523	5	0,5737	3	16,65
	4	0,1098	0,1683	50	-0,0518	-0,029	82	0,0727	0,0572	32	0,2092	6	-6,03
	5	0,1187	0,0532	48	-0,1072	0,2022	26	0,0695	0,0892	33	0,1845	5	-8,16
29	1	-0,6193	-0,0842	72	-0,5831	-0,0426	8	0,0338	0,1414	6	0,229	4	-4,41
	2	0,7319	-1,5136	622	0,0369	-0,392	732	0,0976	1,1238	351	0,5209	2	2,94
	3	0,0153	-0,0431	8	-0,0359	-0,1643	7	0,0261	0,0854	0	0,5393	3	15,29
	4	0,1187	0,0012	10.526	-0,1072	-0,0223	194	0,0695	0,0307	156	0,2693	5	-1,49
	5	0,3107	-0,1662	116	0,1855	0,1848	122	0,1084	0,3166	60	0,1942	5	-7,35
31	1	-0,633	-0,1028	291	-0,5441	0,6639	15	0,0205	0,5257	10	0,3736	3	4,41
	2	0,1187	0,0975	726	-0,1072	-0,0704	92	0,0695	2,E-14	9,E+14	0,4081	4	0,50
	3	-0,1152	0,1281	11	-0,052	-0,0093	12	0,02	0,0246	81	0,5827	4	14,66
	4	0,1187	0,4915	55	-0,1072	0,05	70	0,0695	0,1034	52	0,6135	4	13,33
	5	0,1187	0,1347	422	-0,1072	0,199	270	0,0695	0,8008	31	0,2007	4	-6,82
32	1	0,1187	-0,0658	964	-0,1072	-0,4424	60	0,0695	0,6662	16	0,6191	3	13,50
	2	1,4394	-1,2729	226	-0,234	0,2108	590	0,1866	0,5592	890	0,4637	3	1,78
	3	0,0153	-0,352	84	-0,0359	-0,3989	10	0,0261	0,4608	0	0,5638	3	16,27
	4	0,1187	-0,229	315	-0,1072	0,0669	410	0,0695	0,2828	16	0,53	2	10,70
	5	0,1187	-1,4956	174	-0,1072	0,6916	321	0,0695	0,5757	93	0,5268	3	8,62
41	1	-0,6193	0,135	227	-0,5831	0,5236	3	0,0338	0,1221	2	1,0209	1	22,50
	2	1,4394	-1,1441	133	-0,234	-1,7543	530	0,1866	0,7948	71	0,5903	3	4,19
	3	0,1458	0,0233	147	-0,0197	0,0234	161	0,0322	0,0653	81	0,6747	5	20,22
	4	0,1098	-0,0295	393	-0,0518	0,0746	125	0,0727	0,0738	473	0,2845	3	-0,50
	5	0,0601	-0,0191	263	0,1634	0,0153	229	0,1219	0,0594	387	0,1166	4	-15,51
50	1	0,1187	-0,3432	198	-0,1072	0,7047	75	0,0695	0,8433	4	0,3933	2	5,33
	2	0,7319	-0,5389	172	0,0369	0,2536	188	0,0976	0,2606	205	0,2996	2	-2,59
	3	-0,1152	0,0255	74	-0,052	-0,004	231	0,02	0,0079	90	0,1155	3	-18,61
	4	0,2801	-0,2431	32	-0,1284	0,4009	13	0,075	0,253	7	0,2122	5	-5,78
	5	0,3107	0,0508	113	0,1855	0,047	100	0,1084	0,1719	104	0,1462	7	-11,89

Tabella 7

Abbiamo poi dato come valori iniziali per k_{31} e k_{32} le stime ottime che avevamo ottenuto dall'identificazione eseguita nel paragrafo precedente in cui k_{03} era fissato. Il parametro di degradazione lo abbiamo fatto partire da 0,019 (Tabella 8).

FFL	S	k_{31_iniz}	k_{31_ott}	CV_ k_{31}	k_{32_iniz}	k_{32_ott}	CV_ k_{32}	k_{03_iniz}	k_{03_ott}	CV_ k_{03}	sd_post	nc	AIC
4	1	-0,2298	-0,2298	21	0,3508	0,3508	22	0,019	0,019	141	0,5937	3	12,74
	2	0,1223	0,1223	109	0,1117	0,1117	119	0,019	0,019	469	0,6331	3	4,90
	3	0,0013	0,0013	26	0,0036	0,0036	59	0,019	0,019	90	0,2881	4	0,57
	4	-0,1087	-0,0265	179	0,0398	0,086	76	0,019	0,1082	56	0,2125	2	-5,75
	5	0,0003	0,0003	76	0,1698	0,1698	0	0,019	0,019	107	0,2934	2	-0,74
5	1	-0,1252	-0,127	33	0,1995	0,2024	33	0,019	0,0291	167	0,4442	3	7,52
	2	0,0722	0,0704	14	0,0428	0,0323	67	0,019	0,	1,E+14	0,0702	2	-17,10
	3	-0,061	-0,0757	37	0,0219	0,0352	56	0,019	0,0331	76	0,2678	6	-0,11
	4	-0,016	-0,0416	173	0,0247	0,1391	64	0,019	0,1525	66	0,3218	2	1,72
	5	-0,0261	-0,0261	103	0,2154	0,2154	16	0,019	0,019	89	0,2773	4	-1,64
7	1	0,0013	0,0035	5.600	-0,0292	-0,0314	522	0,019	0,018	517	0,3696	4	4,21
	2	0,0616	0,3432	248	-0,0447	-0,0392	1895	0,019	0,1384	90	0,5114	2	2,76
	3	0,2016	0,2016	6	-0,0262	-0,0262	41	0,019	0,019	5	0,3962	7	8,51
	4	-0,0477	-0,0441	8	-0,0274	-0,0251	14	0,019	0,0151	30	0,107	4	-18,11
	5	0,0013	0,0018	785	0,0315	0,0295	168	0,019	0,0156	200	0,1652	6	-9,93
9	1	-0,2016	-0,2122	13	0,1886	0,1962	24	0,019	0,0266	60	0,5224	2	10,44
	2	0,2613	0,5645	15	0,1382	0,4317	32	0,019	0,1068	20	0,2545	1	-4,22
	3	0,1937	0,1908	19	0,1538	0,1528	59	0,019	0,0187	93	0,6481	2	19,33
	4	0,0013	0,1326	41	-0,5599	-0,5065	15	0,019	0,0068	49	0,535	3	10,87
	5	0,0278	0,0278	136	-0,0802	-0,0802	114	0,019	0,019	280	0,8089	5	15,48
11	1	-0,3257	-0,3265	54	-0,2867	-0,2854	125	0,019	0,0181	508	0,7325	3	16,52
	2	0,2523	0,1411	232	0,0482	0,064	112	0,019	4,E-12	1,E+12	0,3236	3	-1,82
	3	-0,1764	-0,1764	78	-0,0513	-0,0513	56	0,019	0,019	46	0,5726	4	16,61
	4	0,1337	0,1451	32	0,0136	0,0037	180	0,019	0,0311	60	0,2137	6	-5,65
	5	0,0105	0,0212	136	0,0654	0,0839	15	0,019	0,0363	54	0,2192	6	-5,41
29	1	0,0013	0,0192	215	-0,0161	-0,0116	16	0,019	0,	72261	0,1963	4	-7,18
	2	-0,0019	-0,6031	216	-0,014	-0,1135	218	0,019	0,3947	46	0,5546	2	3,57
	3	-0,0257	-0,0257	170	-0,0561	-0,0561	12	0,019	0,019	25	0,6096	3	17,99
	4	-0,0142	-0,0142	170	-0,0135	-0,0135	125	0,019	0,019	158	0,2644	5	-1,82
	5	-0,0019	0,0075	717	0,012	-0,0014	1974	0,019	0,0031	1885	0,2157	5	-5,67
31	1	-0,1439	-0,0949	177	0,0496	0,2936	24	0,019	0,2259	14	0,4224	3	6,61
	2	0,1606	0,0975	726	-0,0778	-0,0704	92	0,019	2,E-14	9,E+14	0,4081	4	0,50
	3	0,1044	0,1044	88	0,0061	0,0061	508	0,019	0,019	280	0,578	4	14,50
	4	0,5621	0,5623	55	-0,1023	-0,103	22	0,019	0,0193	0	0,6589	4	14,62
	5	0,0013	-0,0267	92	-0,0095	0,0233	96	0,019	0,0009	8241	0,2434	5	-3,74
32	1	0,0013	0,0483	576	-0,0191	-0,0161	827	0,019	0,0006	48428	0,669	5	14,89
	2	0,0013	-0,0068	3.063	-0,0106	-0,0106	1486	0,019	3,E-07	4,E+08	0,6424	3	5,04
	3	0,0013	0,0013	3.315	-0,0221	-0,0221	7	0,019	0,019	629	0,7504	3	22,56
	4	-0,208	-0,208	31	0,0727	0,0727	51	0,019	0,019	456	0,4334	3	7,08
	5	0,0013	0,0366	325	-0,0963	-0,1377	38	0,019	0,0211	260	0,5486	5	9,27
41	1	0,0125	0,0125	176	0,4709	0,4709	119	0,019	0,019	466	0,9612	1	21,41
	2	-0,0839	-0,094	1.027	0,062	0,0144	8582	0,019	0,0644	803	0,9818	2	9,28
	3	0,0013	0,0141	138	0,0074	-0,01	139	0,019	0,0074	589	0,6757	4	20,25
	4	-0,0105	-0,0285	320	0,0573	0,0654	65	0,019	0,0706	354	0,2854	3	-0,44
	5	0,0013	0,0007	2.142	0,0441	0,0336	61	0,019	0,0001	47069	0,1015	5	-17,73

50	1	-0,0528	-0,0528	63	0,0645	0,0645	51	0,019	0,019	54	0,3974	3	5,52
	2	-0,3215	-0,3943	188	0,0628	0,2136	192	0,019	0,2238	359	0,3038	2	-2,45
	3	0,0013	0,0399	39	0,0133	-0,0118	75	0,019	0,0081	61	0,1229	3	-17,25
	4	-0,1049	-0,1664	35	0,0794	0,2064	17	0,019	0,1186	8	0,2521	5	-2,68
	5	0,0013	0,0013	1.239	0,0099	0,0099	286	0,019	0,019	465	0,189	3	-7,78

Figura 8

I risultati sono stati confrontati con la tabella ottenuta al passo precedente e abbiamo scelto i fit migliori, ovvero quelli con deviazione standard a posteriori minore (Tabella 9).

FFL	S	k _{31_iniz}	k _{31_ott}	CV_k ₃₁	k _{32_iniz}	k _{32_ott}	CV_k ₃₂	k _{03_iniz}	k _{03_ott}	CV_k ₀₃	sd_post	nc	AIC	
4	x	1	0,1187	0,2504	81	-0,1072	0,1543	90	0,0695	0,1566	37	0,2376	4	-3,75
		2	0,0244	0,2092	101	0,3077	0,3556	83	0,0087	0,1515	120	0,4948	3	2,43
		3	0,0153	-0,0227	265	-0,0359	0,023	542	0,0261	0,0933	602	0,2801	2	0,00
		4	0,4505	0,14	26	-0,2051	0,1978	16	0,0773	0,3163	2	0,1489	5	-12,16
		5	0,1187	0,0351	73	-0,1072	0,2448	14	0,0695	0,0624	22	0,1362	4	-13,03
5	z	1	-0,6056	-0,0741	86	-0,622	0,3375	25	0,0471	0,1532	23	0,1611	5	-10,73
		2	0,1187	0,0704	14	-0,1072	0,0323	67	0,0695	7,E-11	5,E+10	0,0702	2	-17,10
		3	0,1187	0,0917	3	-0,1072	0,1513	0	0,0695	0,2026	4	0,2288	6	-3,57
		4	0,2801	0,2781	23	-0,1284	0,3274	60	0,075	0,4255	54	0,2213	2	-5,02
		5	-0,0261	-0,0261	103	0,2154	0,2154	16	0,019	0,019	89	0,2773	4	-1,64
7	x	1	0,0013	0,0035	5.600	-0,0292	-0,0314	522	0,019	0,018	517	0,3696	4	4,21
		2	0,0244	0,4761	1.163	0,3077	0,0244	3710	0,0087	0,2317	877	0,5055	1	2,64
		3	0,1458	0,185	26	-0,0197	-0,0213	52	0,0322	0,0171	33	0,393	7	8,33
		4	-0,0477	-0,0441	8	-0,0274	-0,0251	14	0,019	0,0151	30	0,107	4	-18,11
		5	0,1187	0,0762	18	-0,1072	-0,1316	24	0,0695	0,2091	47	0,1281	6	-14,00
9	x	1	0,1187	-0,3639	8	-0,1072	0,3502	15	0,0695	0,0592	30	0,416	6	6,34
		2	1,4394	0,5921	30	-0,234	0,4722	44	0,1866	0,1152	40	0,251	3	-4,36
		3	0,1937	0,1908	19	0,1538	0,1528	59	0,019	0,0187	93	0,6481	2	19,33
		4	0,1098	0,1804	13	-0,0518	-0,4833	10	0,0727	0,001	738	0,3815	3	4,78
		5	0,1854	0,0523	0	0,1745	-0,3691	10	0,1151	0,0943	29	0,5151	5	8,26
11	z	1	-0,3257	-0,3265	54	-0,2867	-0,2854	125	0,019	0,0181	508	0,7325	3	16,52
		2	0,1187	0,1413	227	-0,1072	0,0637	109	0,0695	6,E-11	8,E+10	0,3236	3	-1,82
		3	-0,1764	-0,1764	78	-0,0513	-0,0513	56	0,019	0,019	46	0,5726	4	16,61
		4	0,1098	0,1683	50	-0,0518	-0,029	82	0,0727	0,0572	32	0,2092	6	-6,03
		5	0,1187	0,0532	48	-0,1072	0,2022	26	0,0695	0,0892	33	0,1845	5	-8,16
29	z	1	0,0013	0,0192	215	-0,0161	-0,0116	16	0,019	3,E-05	72261	0,1963	4	-7,18
		2	0,7319	-1,5136	622	0,0369	-0,392	732	0,0976	1,1238	351	0,5209	2	2,94
		3	0,0153	-0,0431	8	-0,0359	-0,1643	7	0,0261	0,0854	0	0,5393	3	15,29
		4	-0,0142	-0,0142	170	-0,0135	-0,0135	125	0,019	0,019	158	0,2644	5	-1,82
		5	0,3107	-0,1662	116	0,1855	0,1848	122	0,1084	0,3166	60	0,1942	5	-7,35
31	x	1	-0,633	-0,1028	291	-0,5441	0,6639	15	0,0205	0,5257	10	0,3736	3	4,41
		2	0,1606	0,0975	726	-0,0778	-0,0704	92	0,019	2,E-14	9,E+14	0,4081	4	0,50

x	3	0,1044	0,1044	88	0,0061	0,0061	508	0,019	0,019	280	0,578	4	14,50	
	4	0,1187	0,4915	55	-0,1072	0,05	70	0,0695	0,1034	52	0,6135	4	13,33	
	5	0,1187	0,1347	422	-0,1072	0,199	270	0,0695	0,8008	31	0,2007	4	-6,82	
32	x	1	0,1187	-0,0658	964	-0,1072	-0,4424	60	0,0695	0,6662	16	0,6191	3	13,50
	x	2	1,4394	-1,2729	226	-0,234	0,2108	590	0,1866	0,5592	890	0,4637	3	1,78
	x	3	0,0153	-0,352	84	-0,0359	-0,3989	10	0,0261	0,4608	0	0,5638	3	16,27
	x	4	-0,208	-0,208	31	0,0727	0,0727	51	0,019	0,019	456	0,4334	3	7,08
	x	5	0,1187	-1,4956	174	-0,1072	0,6916	321	0,0695	0,5757	93	0,5268	3	8,62
41	x	1	0,0125	0,0125	176	0,4709	0,4709	119	0,019	0,019	466	0,9612	1	21,41
	x	2	1,4394	-1,1441	133	-0,234	-1,7543	530	0,1866	0,7948	71	0,5903	3	4,19
	x	3	0,1458	0,0233	147	-0,0197	0,0234	161	0,0322	0,0653	81	0,6747	5	20,22
	x	4	0,1098	-0,0295	393	-0,0518	0,0746	125	0,0727	0,0738	473	0,2845	3	-0,50
	x	5	0,0013	0,0007	2.142	0,0441	0,0336	61	0,019	0,0001	47069	0,1015	5	-17,73
50	x	1	0,1187	-0,3432	198	-0,1072	0,7047	75	0,0695	0,8433	4	0,3933	2	5,33
	x	2	0,7319	-0,5389	172	0,0369	0,2536	188	0,0976	0,2606	205	0,2996	2	-2,59
	x	3	-0,1152	0,0255	74	-0,052	-0,004	231	0,02	0,0079	90	0,1155	3	-18,61
	x	4	0,2801	-0,2431	32	-0,1284	0,4009	13	0,075	0,253	7	0,2122	5	-5,78
	x	5	0,3107	0,0508	113	0,1855	0,047	100	0,1084	0,1719	104	0,1462	7	-11,89

Tabella 9

Da qui è iniziata una serie di nuove identificazioni partendo da stime iniziali ancora differenti dalle precedenti per cercare di migliorare tutti quei risultati con CV considerati elevati o non validi.

In alcuni casi i CV del parametro k_{03} risultano molto alti, dall'ordine di 1×10^{10} a ordini superiori, a fronte di valori dei parametri vicini allo zero perciò ripetiamo la stima parametrica fissando il parametro k_{03} a zero.

Così facendo si abbassa la deviazione standard a posteriori quindi il fit migliora. Nel caso della serie temporale numero 5 del FFL 41, dopo aver fissato a zero il parametro di degradazione, si trova che il fit è migliorato (la deviazione standard a posteriori passa da -17,73 a -19,74) ma il parametro k_{31} viene stimato a circa 0,0008 con un coefficiente di variazione alto (circa 1155 %). Allora fissiamo a zero sia k_{03} sia k_{31} e, in questo modo, si ottiene una buona stima per k_{32} (0,033 con CV=45%) e il fit migliora ulteriormente (sd_post=-21,74).

A questo punto ci siamo occupati dei 25 casi in cui i CV di parametri con valori diversi da zero sono alti. Abbiamo provato a ripetere l'identificazione dando come valori iniziali le medie dei parametri calcolate considerando tutti gli altri casi in cui le stime sono accettabili. I valori medi sono: $k_{31} = 0,0782$, $k_{32} = 0,1393$ e $k_{03} = 0,1092$. La situazione che ne risulta è la seguente:

-In 3 casi il fit migliora quindi prendiamo i nuovi risultati.

-In 2 caso il fit migliora ma ho ancora CV alti o a zero; decidiamo di tenere quest'ultimo risultato perché comunque il fit è migliorato.

-In 4 casi il fit peggiora di poco ma i CV diventano accettabili per cui scegliamo di tenere questo risultato.

-nei restanti 16 casi il fit peggiora.

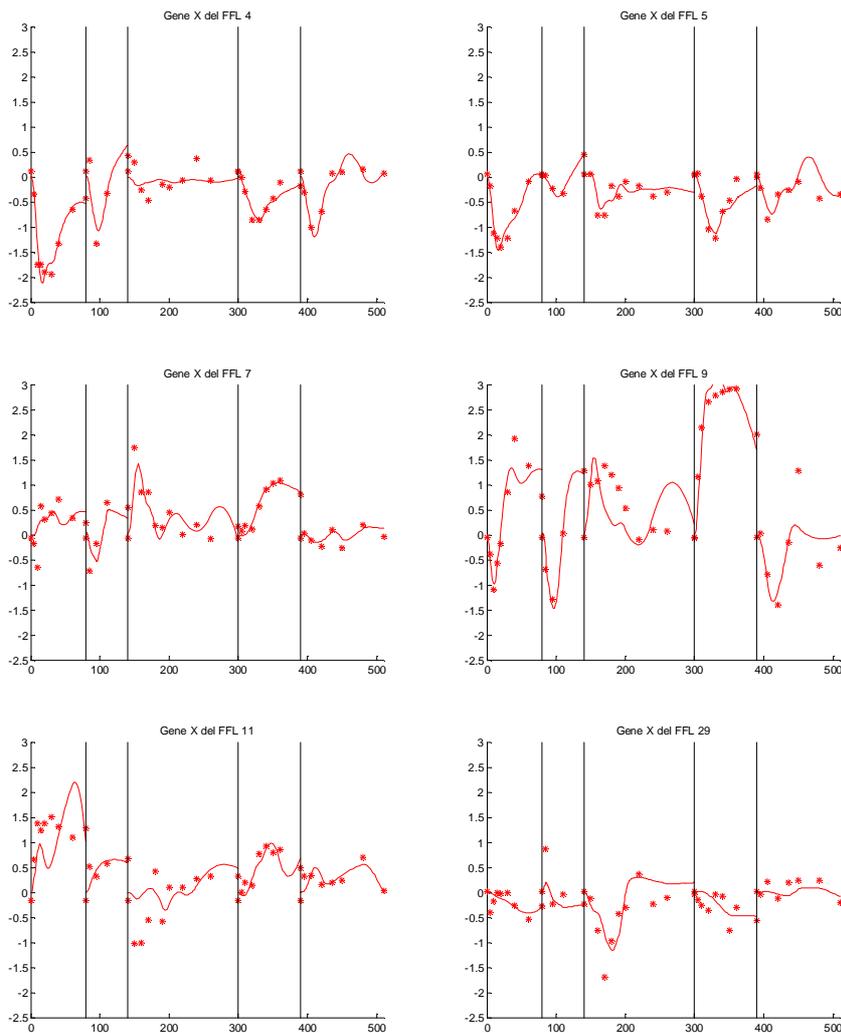
La Tabella 10 riassume i risultati ottenuti.

FFL	S	k_{31_iniz}	k_{31_ott}	$CV_{k_{31}}$	k_{32_iniz}	k_{32_ott}	$CV_{k_{32}}$	k_{03_iniz}	k_{03_ott}	$CV_{k_{03}}$	sd_post	nc	AIC
4	1	0,1187	0,2504	81	-0,1072	0,1543	90	0,0695	0,1566	37	0,2376	4	-3,75
	2	0,0244	0,2092	101	0,3077	0,3556	83	0,0087	0,1515	120	0,4948	3	2,43
	3	0,0153	-0,0227	265	-0,0359	0,023	542	0,0261	0,0933	602	0,2801	2	0,00
	4	0,4505	0,14	26	-0,2051	0,1978	16	0,0773	0,3163	2	0,1489	5	-12,16
	5	0,1187	0,0351	73	-0,1072	0,2448	14	0,0695	0,0624	22	0,1362	4	-13,03
5	1	-0,6056	-0,0741	86	-0,622	0,3375	25	0,0471	0,1532	23	0,1611	5	-10,73
	2	0,1187	0,0704	10	-0,1072	0,0323	16		0,		0,0573	2	-19,10
	3	0,1187	0,0917	3	-0,1072	0,1513	0	0,0695	0,2026	4	0,2288	6	-3,57
	4	0,2801	0,2781	23	-0,1284	0,3274	60	0,075	0,4255	54	0,2213	2	-5,02
	5	-0,0261	-0,0261	103	0,2154	0,2154	16	0,019	0,019	89	0,2773	4	-1,64
7	1	0,0782	-0,2301	132	0,1393	0,1698	140	0,1092	0,1289	157	0,401	4	5,68
	2	0,0782	0,892	78	0,1393	0,4079	268	0,1092	0,7174	8	0,4886	1	2,30
	3	0,1458	0,185	26	-0,0197	-0,0213	52	0,0322	0,0171	33	0,393	7	8,33
	4	-0,0477	-0,0441	8	-0,0274	-0,0251	14	0,019	0,0151	30	0,107	4	-18,11
	5	0,1187	0,0762	18	-0,1072	-0,1316	24	0,0695	0,2091	47	0,1281	6	-14,00
9	1	0,1187	-0,3639	8	-0,1072	0,3502	15	0,0695	0,0592	30	0,416	6	6,34
	2	1,4394	0,5921	30	-0,234	0,4722	44	0,1866	0,1152	40	0,251	3	-4,36
	3	0,1937	0,1908	19	0,1538	0,1528	59	0,019	0,0187	93	0,6481	2	19,33
	4	0,1098	0,1804	13	-0,0518	-0,4833	10	0,0727	0,001	738	0,3815	3	4,78
	5	0,0782	0,0953	94	0,1393	-0,1911	116	0,1092	0,0818	111	0,5878	4	10,37
11	1	-0,3257	-0,3265	54	-0,2867	-0,2854	125	0,019	0,0181	508	0,7325	3	16,52
	2	0,1187	0,1413	34	-0,1072	0,0636	72		0,		0,2643	3	-3,82
	3	-0,1764	-0,1764	78	-0,0513	-0,0513	56	0,019	0,019	46	0,5726	4	16,61
	4	0,1098	0,1683	50	-0,0518	-0,029	82	0,0727	0,0572	32	0,2092	6	-6,03
	5	0,1187	0,0532	48	-0,1072	0,2022	26	0,0695	0,0892	33	0,1845	5	-8,16
29	1	0,0013	0,0191	122	-0,0161	-0,0116	21		0,		0,1817	4	-9,18
	2	0,7319	-1,5136	622	0,0369	-0,392	732	0,0976	1,1238	351	0,5209	2	2,94
	3	0,0782	-0,5895	3	0,1393	-0,1513	0	0,1092	0,2777	0	0,4367	5	10,65
	4	-0,0142	-0,0142	170	-0,0135	-0,0135	125	0,019	0,019	158	0,2644	5	-1,82
	5	0,3107	-0,1662	116	0,1855	0,1848	122	0,1084	0,3166	60	0,1942	5	-7,35
31	1	-0,633	-0,1028	291	-0,5441	0,6639	15	0,0205	0,5257	10	0,3736	3	4,41
	2	0,1606	0,0975	82	-0,0778	-0,0704	39		0,		0,3332	4	-1,50
	3	0,1044	0,1044	88	0,0061	0,0061	508	0,019	0,019	280	0,578	4	14,50
	4	0,1187	0,4915	55	-0,1072	0,05	70	0,0695	0,1034	52	0,6135	4	13,33
	5	0,1187	0,1347	422	-0,1072	0,199	270	0,0695	0,8008	31	0,2007	4	-6,82
32	1	0,1187	-0,0658	964	-0,1072	-0,4424	60	0,0695	0,6662	16	0,6191	3	13,50
	2	1,4394	-1,2729	226	-0,234	0,2108	590	0,1866	0,5592	890	0,4637	3	1,78
	3	0,0153	-0,352	84	-0,0359	-0,3989	10	0,0261	0,4608	0	0,5638	3	16,27
	4	0,0782	-0,3355	19	0,1393	0,1493	45	0,1092	0,2088	30	0,4821	2	8,99
	5	0,0782	-1,0418	61	0,1393	0,4011	192	0,1092	0,4343	16	0,5239	3	8,53

41	1	0,0782	0,2185	37	0,1393	0,373	17		0		0,4681	5	7,85
	2	1,4394	-1,1441	133	-0,234	-1,7543	530	0,1866	0,7948	71	0,5903	3	4,19
	3	0,0782	0,0307	102	0,1393	0,0506	11	0,1092	0,1563	16	0,6836	5	20,51
	4	0,1098	-0,0295	393	-0,0518	0,0746	125	0,0727	0,0738	473	0,2845	3	-0,50
	5		0,		0,0441	0,033	45		0,		0,0857	5	-21,74
50	1	0,0782	-0,0878	106	0,1393	0,1473	82	0,1092	0,1394	103	0,4022	4	5,73
	2	0,7319	-0,5389	172	0,0369	0,2536	188	0,0976	0,2606	205	0,2996	2	-2,59
	3	-0,1152	0,0255	74	-0,052	-0,004	231	0,02	0,0079	90	0,1155	3	-18,61
	4	0,2801	-0,2431	32	-0,1284	0,4009	13	0,075	0,253	7	0,2122	5	-5,78
	5	0,3107	0,0508	113	0,1855	0,047	100	0,1084	0,1719	104	0,1462	7	-11,89

Tabella 10

I plot del fit della predizione contro i dati sono mostrati nella Figura 4.7.



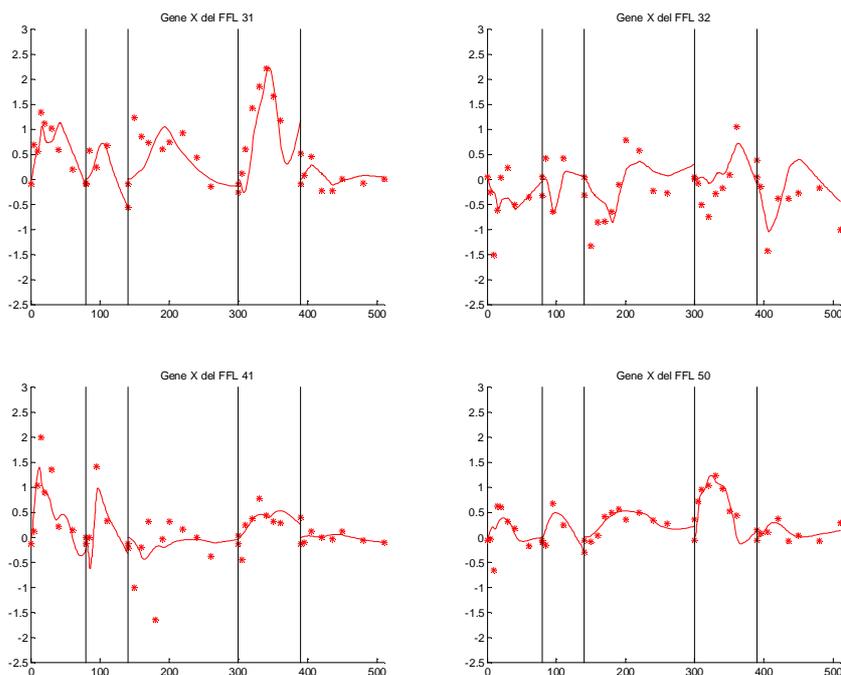


Figura 4.7

4.2. Modello “feedback-integrativo”

4.2.1. Stima di a_i e b_i .

Stimiamo i parametri a_i e b_i del modello usando lo stimatore ai minimi quadrati non lineari pesati. Il parametro di degradazione δ_i viene fissato al valore 0,019 e partiamo dai valori iniziali $a_i = 0,009$ e $b_i = 0,015$.

I tre valori numerici corrispondono ai valori medi ottenuti nel lavoro di Altafini [3] fittando il modello con i dati di espressione e i dati proteici.

Applichiamo la procedura di ottimizzazione e calcoliamo i vari indici che ci servono.

I risultati sono riportati nelle Tabella 11.

FFL	S	ai_iniz	ai_ott	CV_ai	bi_iniz	bi_ott	CV_bi	sd_post	nc	AIC
4	1	0,009	0,1764	16	0,015	-0,1128	14	0,4737	6	8,06
	2	0,009	0,8128	38	0,015	-0,0967	56	0,5168	4	2,89
	3	0,009	0,1616	41	0,015	-0,015	47	0,2145	7	-5,99
	4	0,009	0,1671	24	0,015	-0,0384	20	0,2318	7	-4,80
	5	0,009	0,4672	14	0,015	-0,0647	23	0,2488	6	-3,93
5	1	0,009	0,2511	17	0,015	-0,082	15	0,3302	6	1,57
	2	0,009	3,E-13	4,E+14	0,015	0,003	535	0,3389	3	-1,33

	3	0,009	0,051	50	0,015	-0,0158	36	0,2794	8	0,12
	4	0,009	0,2367	21	0,015	-0,0569	20	0,282	7	-1,27
	5	0,009	0,034	118	0,015	-0,0141	55	0,3132	4	-0,24
7	1	0,009	0,0858	98	0,015	0,0202	54	0,3659	5	3,42
	2	0,009	2,E-10	2,E+11	0,015	0,0144	173	0,528	2	3,11
	3	0,009	0,3108	24	0,015	0,0585	38	0,5281	9	14,12
	4	0,009	0,0248	77	0,015	0,0301	15	0,185	4	-8,86
	5	0,009	0,2412	56	0,015	-0,0095	71	0,1464	7	-12,41
9	1	0,009	8,E-14	1,E+14	0,015	0,0275	82	0,8448	4	18,48
	2	0,009	2,E-12	5,E+13	0,015	0,0134	376	1,0724	1	10,19
	3	0,009	0,1531	15	0,015	0,0634	15	0,3239	8	3,37
	4	0,009	0,0747	17	0,015	0,1401	9	0,4765	5	8,17
	5	0,009	0,5448	13	0,015	-0,1141	24	0,4384	6	5,14
11	1	0,009	0,0948	25	0,015	0,0792	15	0,3948	6	4,79
	2	0,009	0,0814	94	0,015	0,0313	38	0,2384	3	-4,84
	3	0,009	0,3456	25	0,015	-0,0473	42	0,4616	6	11,16
	4	0,009	0,0571	35	0,015	0,0315	15	0,1796	5	-9,39
	5	0,009	0,0371	93	0,015	0,0129	47	0,2421	5	-4,36
29	1	0,009	2,E-12	4,E+12	0,015	-0,0087	58	0,1913	6	-8,26
	2	0,009	0,9002	137	0,015	0,0221	251	0,5186	2	2,93
	3	0,009	0,2763	17	0,015	-0,0668	24	0,3962	4	7,80
	4	0,009	3,E-14	2,E+14	0,015	-0,0121	42	0,2209	6	-5,67
	5	0,009	0,08	75	0,015	0,0076	62	0,1654	7	-10,46
31	1	0,009	0,239	18	0,015	0,0711	16	0,3107	6	0,47
	2	0,009	0,497	51	0,015	0,0568	37	0,3096	4	-2,23
	3	0,009	0,078	25	0,015	0,0427	22	0,3638	7	4,57
	4	0,009	0,1086	12	0,015	0,0898	8	0,2531	5	-3,22
	5	0,009	0,051	398	0,015	-0,0023	258	0,2288	6	-5,26
32	1	0,009	0,1008	133	0,015	-0,0215	84	0,593	6	12,11
	2	0,009	0,4473	953	0,015	0,0045	743	0,5342	4	3,22
	3	0,009	0,4295	9	0,015	-0,0924	13	0,2605	8	-1,43
	4	0,009	2,E-11	1,E+12	0,015	0,0047	264	0,542	3	10,49
	5	0,009	0,0001	47.148	0,015	-0,0149	83	0,5645	5	9,19
41	1	0,009	0,5142	22	0,015	0,1078	26	0,5559	7	10,94
	2	0,009	0,9359	18	0,015	0,1227	34	0,3952	3	0,21
	3	0,009	0,3516	35	0,015	-0,0433	59	0,5809	7	16,22
	4	0,009	0,0666	78	0,015	0,0195	38	0,2738	5	-1,80
	5	0,009	0,161	153	0,015	0,0028	139	0,1025	4	-18,12
50	1	0,009	0,371	81	0,015	0,0233	70	0,3764	6	3,93
	2	0,009	0,6616	60	0,015	0,0466	56	0,2535	3	-4,23
	3	0,009	0,0219	43	0,015	0,0139	17	0,1411	7	-14,91
	4	0,009	0,1737	21	0,015	0,0625	18	0,3245	6	1,26
	5	0,009	5,E-11	2,E+11	0,015	0,0028	145	0,1829	5	-8,85

Tabella 11

Dalla Tabella 11 possiamo notare che ci sono dei casi in cui i coefficienti di variazione assumono valori superiori al 100%.

Decidiamo quindi di utilizzare alcune strategie per cercare di ottenere dei fit migliori. Verranno spiegati di seguito i vari passi che abbiamo eseguito per ottenere una stima parametrica il più possibile soddisfacente.

Nella Tabella 11 notiamo che ci sono dei casi in cui la stima ottima del parametro a_i assume valori molto bassi, pressoché pari a zero, e il corrispondente CV supera di molto il 100% (da un ordine di grandezza di 1×10^4 ad ordini superiori fino a 1×10^{14}). I casi in questione sono i seguenti:

FFL 5, serie 2

FFL 7, serie 2

FFL 9, serie 1 e 2

FFL 29, serie 1 e 4

FFL 32, serie 4 e 5

FFL 50, serie 5

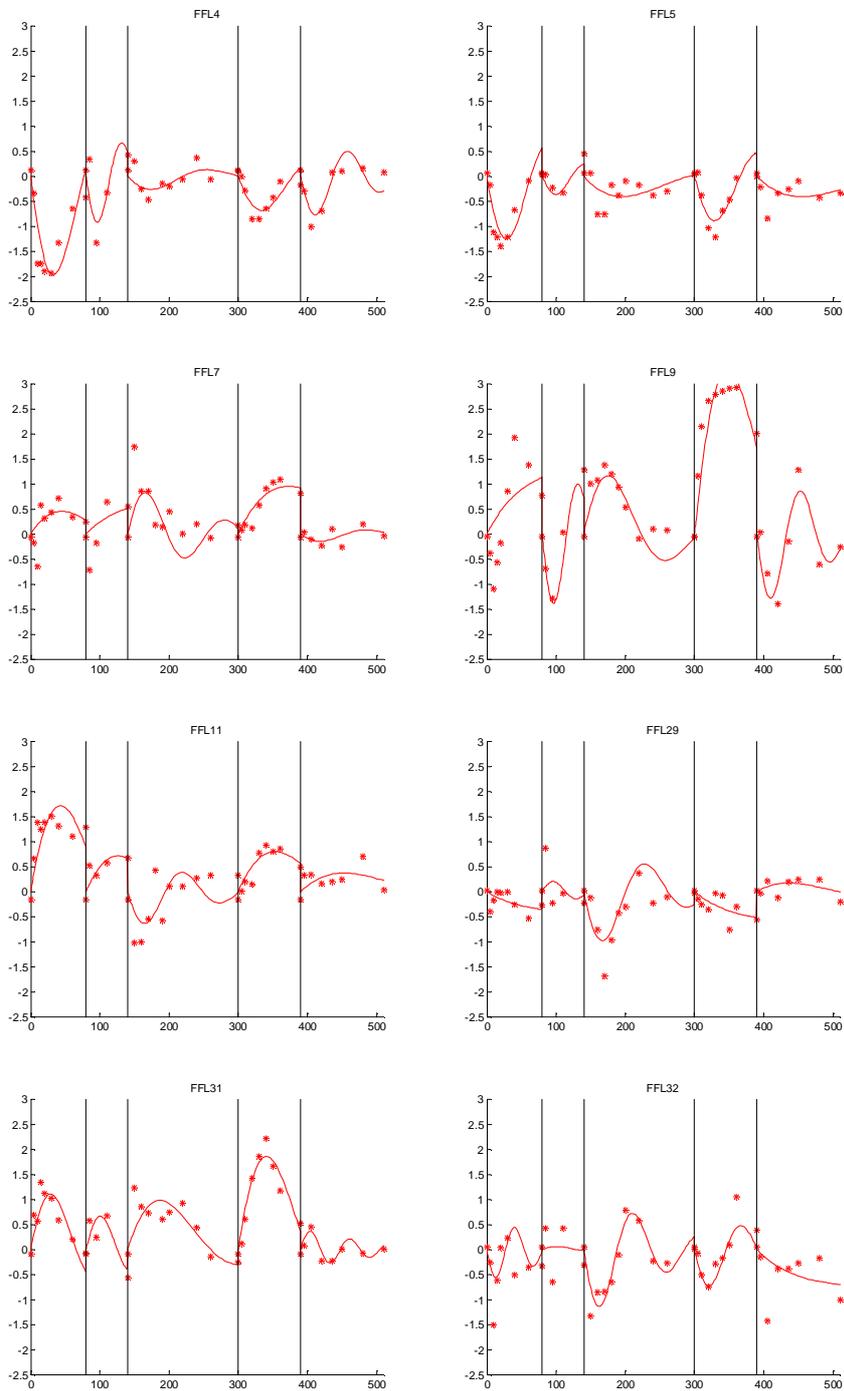
Abbiamo allora ripetuto l'identificazione, solo per i suddetti casi, fissando il parametro a_i a zero e quindi stimando solamente il parametro b_i . Confrontando i coefficienti di Akaike ottenuti in precedenza con quelli ottenuti ora, per i medesimi casi in questione, si vede che i secondi sono più bassi. Ciò significa che il modello privato del termine in retroazione $a_i p_i$ descrive meglio i dati sperimentali (Tabella 12).

FFL	S	ai_iniz	ai_ott	CV_ai	bi_iniz	bi_ott	CV_bi	sd_post	nc	AIC
4	1	0,009	0,1764	16	0,015	-0,1128	14	0,4737	6	8,06
	2	0,009	0,8128	38	0,015	-0,0967	56	0,5168	4	2,89
	3	0,009	0,1616	41	0,015	-0,015	47	0,2145	7	-5,99
	4	0,009	0,1671	24	0,015	-0,0384	20	0,2318	7	-4,80
	5	0,009	0,4672	14	0,015	-0,0647	23	0,2488	6	-3,93
5	1	0,009	0,2511	17	0,015	-0,082	15	0,3302	6	1,57
	2		0		0,015	0,003	221	0,2935	3	-3,33
	3	0,009	0,051	50	0,015	-0,0158	36	0,2794	8	0,12
	4	0,009	0,2367	21	0,015	-0,0569	20	0,282	7	-1,27
	5	0,009	0,034	118	0,015	-0,0141	55	0,3132	4	-0,24
7	1	0,009	0,0858	98	0,015	0,0202	54	0,3659	5	3,42
	2		0		0,015	0,0144	71	0,4572	2	1,11
	3	0,009	0,3108	24	0,015	0,0585	38	0,5281	9	14,12
	4	0,009	0,0248	77	0,015	0,0301	15	0,185	4	-8,86
	5	0,009	0,2412	56	0,015	-0,0095	71	0,1464	7	-12,41
9	1		0		0,015	0,0275	41	0,7902	4	16,48
	2		0		0,015	0,0134	155	0,9287	1	8,19
	3	0,009	0,1531	15	0,015	0,0634	15	0,3239	8	3,37
	4	0,009	0,0747	17	0,015	0,1401	9	0,4765	5	8,17
	5	0,009	0,5448	13	0,015	-0,1141	24	0,4384	6	5,14
11	1	0,009	0,0948	25	0,015	0,0792	15	0,3948	6	4,79

FFL	S	ai_iniz	ai_ott	CV_ai	bi_iniz	bi_ott	CV_bi	sd_post	nc	AIC
4	1	0,009	0,1764	16	0,015	-0,1128	14	0,4737	6	8,06
	2	0,009	0,8128	38	0,015	-0,0967	56	0,5168	4	2,89
	3	0,009	0,1616	41	0,015	-0,015	47	0,2145	7	-5,99
	4	0,009	0,1671	24	0,015	-0,0384	20	0,2318	7	-4,80
	5	0,009	0,4672	14	0,015	-0,0647	23	0,2488	6	-3,93
5	1	0,009	0,2511	17	0,015	-0,082	15	0,3302	6	1,57
	2	0,2117	0,5491	60	0,0058	-0,0324	43	0,1791	4	-7,70
	3	0,009	0,051	50	0,015	-0,0158	36	0,2794	8	0,12
	4	0,009	0,2367	21	0,015	-0,0569	20	0,282	7	-1,27
	5	0,009	0,034	118	0,015	-0,0141	55	0,3132	4	-0,24
7	1	0,009	0,0858	98	0,015	0,0202	54	0,3659	5	3,42
	2		0		0,015	0,0144	71	0,4572	2	1,11
	3	0,009	0,3108	24	0,015	0,0585	38	0,5281	9	14,12
	4	0,009	0,0248	77	0,015	0,0301	15	0,185	4	-8,86
	5	0,009	0,2412	56	0,015	-0,0095	71	0,1464	7	-12,41
9	1		0		0,015	0,0275	41	0,7902	4	16,48
	2	0,2117	0,8079	23	0,0058	-0,1456	35	0,4652	3	1,84
	3	0,009	0,1531	15	0,015	0,0634	15	0,3239	8	3,37
	4	0,009	0,0747	17	0,015	0,1401	9	0,4765	5	8,17
	5	0,009	0,5448	13	0,015	-0,1141	24	0,4384	6	5,14
11	1	0,009	0,0948	25	0,015	0,0792	15	0,3948	6	4,79
	2	0,009	0,0814	94	0,015	0,0313	38	0,2384	3	-4,84
	3	0,009	0,3456	25	0,015	-0,0473	42	0,4616	6	11,16
	4	0,009	0,0571	35	0,015	0,0315	15	0,1796	5	-9,39
	5	0,009	0,0371	93	0,015	0,0129	47	0,2421	5	-4,36
29	1		0		0,015	-0,0087	29	0,179	6	-10,26
	2	0,009	0,9002	137	0,015	0,0221	251	0,5186	2	2,93
	3	0,009	0,2763	17	0,015	-0,0668	24	0,3962	4	7,80
	4		0		0,015	-0,0121	22	0,2066	6	-7,67
	5	0,009	0,08	75	0,015	0,0076	62	0,1654	7	-10,46
31	1	0,009	0,239	18	0,015	0,0711	16	0,3107	6	0,47
	2	0,009	0,497	51	0,015	0,0568	37	0,3096	4	-2,23
	3	0,009	0,078	25	0,015	0,0427	22	0,3638	7	4,57
	4	0,009	0,1086	12	0,015	0,0898	8	0,2531	5	-3,22
	5	0,2117	1,2059	8	0,0058	0,0447	23	0,1143	6	-16,37
32	1	0,2117	1,3569	29	0,0058	-0,074	53	0,5534	5	10,86
	2	0,009	0,4473	953	0,015	0,0045	743	0,5342	4	3,22
	3	0,009	0,4295	9	0,015	-0,0924	13	0,2605	8	-1,43
	4	0,2117	0,4807	16	0,0058	-0,0619	26	0,3068	8	0,24
	5		0		0,015	-0,0149	42	0,5226	5	7,19
41	1	0,009	0,5142	22	0,015	0,1078	26	0,5559	7	10,94
	2	0,009	0,9359	18	0,015	0,1227	34	0,3952	3	0,21
	3	0,009	0,3516	35	0,015	-0,0433	59	0,5809	7	16,22
	4	0,009	0,0666	78	0,015	0,0195	38	0,2738	5	-1,80
	5	0,009	0,161	153	0,015	0,0028	139	0,1025	4	-18,12
50	1	0,009	0,371	81	0,015	0,0233	70	0,3764	6	3,93
	2	0,009	0,6616	60	0,015	0,0466	56	0,2535	3	-4,23
	3	0,009	0,0219	43	0,015	0,0139	17	0,1411	7	-14,91
	4	0,009	0,1737	21	0,015	0,0625	18	0,3245	6	1,26
	5	0,2117	0,4064	26	0,0058	0,0201	40	0,1378	5	-13,37

Tabella 13

I plot relativi alla tabella sono riportati nella Figura 4.8.



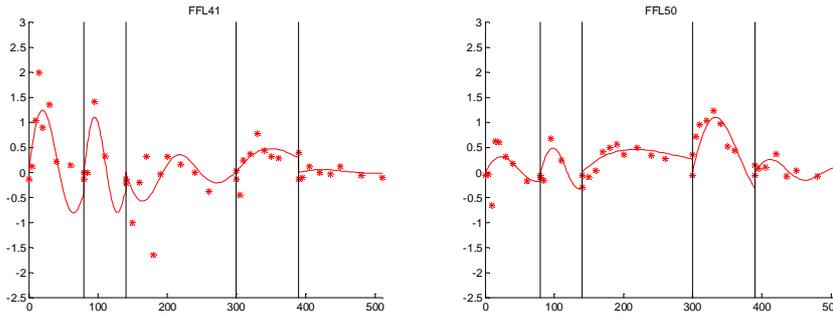


Figura 4.8

4.2.2. Stima di a_i , b_i e δ_i .

Procediamo ora all’identificazione del modello proposto da Altafini, cercando di stimare tutti e tre i parametri a_i , b_i e δ_i .

Per prima cosa applichiamo lo stimatore partendo dai seguenti valori iniziali: $a_i = 0,09$, $b_i = 0,015$, $\delta_i = 0,019$.

Per alcune delle 50 identificazioni il parametro a_i o il parametro δ_i o entrambi raggiungono una stima molto vicina allo zero (dell’ordine di 1×10^{-10} o inferiore) e i corrispondenti coefficienti di variazione (CV) sono molto alti (dell’ordine di 1×10^{10} o superiore). In questi casi fissiamo il parametro o i parametri in questione al valore zero e ripetiamo l’identificazione.

Per la precisione, nei casi in cui sia a_i che δ_i sono quasi nulli, fissiamo a zero prima a_i , poi δ_i . A volte, così facendo, in una delle due identificazioni così costruite, il parametro che viene stimato si “riassesta” nel senso che assume valore diverso da zero con CV accettabili. Altre volte succede che, in entrambe le situazioni (a_i fisso e poi δ_i fisso), rispettivamente δ_i e a_i (cioè i parametri lasciati “liberi” per la stima) continuano ad assumere valori quasi nulli con CV elevati per cui abbiamo necessariamente fissato a zero, contemporaneamente, entrambi. La tabella con i risultati è la Tabella 14.

FFL	S	a_i _iniz	a_i _ott	CV_ a_i	b_i _iniz	b_i _ott	CV_ b_i	δ_i _iniz	δ_i _ott	CV_ δ_i	sd_post	nc	AIC
4	1	0,009	0,2269	19	0,015	-0,2106	21	0,019	0,0748	32	0,2817	5	-0,68
	2	0,009	0,9111	21	0,015	-0,091	52		0,		0,507	1	2,70
	3	0,009	0,1573	40	0,015	-0,0136	81	0,019	0,0141	207	0,2284	6	-4,07
	4	0,009	0,1986	29	0,015	-0,0629	39	0,019	0,0512	61	0,2065	3	-6,26
	5	0,009	0,5261	17	0,015	-0,1135	23	0,019	0,0609	36	0,1547	4	-10,98
5	1	0,009	0,328	17	0,015	-0,1409	18	0,019	0,0621	30	0,2014	3	-6,71
	2		0,		0,015	0,003	138		0,		0,283	2	-3,69
	3	0,009	0,1087	103	0,015	-0,0484	116	0,019	0,0856	135	0,2641	4	-0,42
	4	0,009	0,2442	24	0,015	-0,0759	34	0,019	0,0393	61	0,2707	3	-1,39

	5	0,009	0,1712	139	0,015	-0,1179	140	0,019	0,2247	161	0,2405	3	-3,92
7	1	0,009	0,1112	53	0,015	0,0166	49		0,		0,3583	3	3,04
	2		0,		0,015	0,0097	67		0,		0,4477	2	0,90
	3	0,009	2,	195	0,015	1,0057	215	0,019	0,4575	186	0,1961	9	-6,97
	4	0,009	0,0543	21	0,015	0,0235	11		0,		0,1502	2	-12,61
	5	0,009	0,2593	30	0,015	-0,0071	64		0,		0,1433	6	-12,75
9	1	0,009	0,0499	163	0,015	0,0247	69		0,		0,818	2	17,90
	2		0,		0,015	0,0117	110		0,		0,8888	1	7,75
	3	0,009	0,2035	12	0,015	0,1251	15	0,019	0,0576	20	0,1535	6	-12,35
	4	0,009	0,057	19	0,015	0,2967	10	0,019	0,0858	15	0,1726	4	-9,50
	5	0,009	0,0119	1929	0,015	-0,0028	437		0,		0,8701	4	16,11
11	1	0,009	0,0586	58	0,015	0,2342	23	0,019	0,1549	31	0,1568	5	-11,23
	2		0,		0,015	0,5027	808	0,019	0,96	818	0,1733	3	-8,03
	3	0,009	0,3436	19	0,015	-0,0249	59		0,		0,5195	3	13,76
	4	0,009	0,0778	16	0,015	0,0239	12		0,		0,1568	4	-11,83
	5	0,009	0,0501	43	0,015	0,0092	41		0,		0,2349	4	-4,84
29	1	0,009	0,0356	187	0,015	-0,0069	56		0,		0,189	4	-8,47
	2	0,009	2,	268	0,015	0,1685	342	0,019	0,2515	425	0,5258	1	3,04
	3	0,009	0,2319	29	0,015	-0,0844	43	0,019	0,0406	63	0,3841	6	7,82
	4		0,		0,015	-0,011	76	0,019	0,0151	189	0,2206	5	-5,69
	5	0,009	0,0841	36	0,015	0,0059	45		0,		0,1494	5	-12,08
31	1	0,009	0,3328	20	0,015	0,1367	22	0,019	0,075	33	0,1929	8	-7,49
	2	0,009	0,4549	36	0,015	0,0433	30		0,		0,2845	4	-3,08
	3	0,009	0,8896	1043	0,015	0,8096	1058	0,019	0,6393	1016	0,3032	4	1,59
	4	0,009	0,1083	14	0,015	0,0951	23	0,019	0,0227	62	0,2714	3	-1,35
	5	0,009	0,0674	163	0,015	-0,002	190		0,		0,227	5	-5,39
32	1	0,009	1,5597	111	0,015	-0,2414	149	0,019	0,2038	176	0,5101	3	10,01
	2	0,009	0,4611	458	0,015	0,0065	371		0,		0,5293	3	3,13
	3	0,009	0,4453	10	0,015	-0,1073	21	0,019	0,027	36	0,2615	7	-0,63
	4		0,		0,015	0,0038	99		0,		0,4924	2	7,96
	5	0,009	0,0504	424	0,015	-0,138	234	0,019	0,2139	266	0,515	3	8,26
41	1	0,009	0,4258	39	0,015	0,1695	43	0,019	0,0758	67	0,4674	5	8,44
	2	0,009	0,3702	44	0,015	0,0497	49		0,		0,5504	3	3,52
	3	0,009	1,6002	1.917	0,015	-0,4457	2088	0,019	0,4588	1897	0,6076	5	17,91
	4	0,009	0,0571	154	0,015	0,0259	119	0,019	0,0395	240	0,2925	3	0,00
	5	0,009	0,1566	92	0,015	0,0022	99		0,		0,0983	4	-18,78
50	1	0,009	0,3875	77	0,015	0,0202	121	0,019	0,0081	747	0,4057	3	5,89
	2	0,009	0,707	77	0,015	0,0452	84	0,019	0,0128	475	0,3094	3	-2,26
	3	0,009	0,0209	63	0,015	0,0148	51	0,019	0,0223	111	0,1493	5	-12,96
	4	0,009	0,2551	12	0,015	0,1551	14	0,019	0,0921	19	0,1181	5	-16,33
	5		0,		0,015	0,0329	491	0,019	0,309	511	0,173	3	-9,74

Tabella 14

Ora ripetiamo l'identificazione, per tutti e 50 i casi, dando come valori iniziali le stime ottime che erano state ottenute alla fine del precedente paragrafo ovvero dopo aver stimato il modello "a due parametri" (cioè con δ_i fisso a 0,019).

Così facendo, il fit non può che migliorare visto che è aumentato l'ordine del modello. Fissiamo a zero il parametro a_i nei casi in cui ciò era stato richiesto dalla procedura seguita nel paragrafo precedente e fissiamo a zero il parametro δ_i quando assume valori molto vicini allo zero con CV molto alti.

Il risultato di queste operazioni porta alle stime di Tabella 15.

FFL	S	a_{i_iniz}	a_{i_ott}	CV_ a_i	b_{i_iniz}	b_{i_ott}	CV_ b_i	δ_{i_iniz}	δ_{i_ott}	CV_ δ_i	sd_post	nc	AIC
4	1	0,1764	0,2269	19	-0,1128	-0,2106	21	0,019	0,0748	32	0,2817	5	-0,68
	2	0,8128	0,9111	21	-0,0967	-0,091	52		0,		0,507	1	2,70
	3	0,1616	0,1574	41	-0,015	-0,0136	81	0,019	0,0142	207	0,2284	6	-4,07
	4	0,1671	0,1986	29	-0,0384	-0,0629	39	0,019	0,0512	61	0,2065	3	-6,26
	5	0,4672	0,5261	17	-0,0647	-0,1135	23	0,019	0,0609	36	0,1547	4	-10,98
5	1	0,2511	0,328	17	-0,082	-0,1409	18	0,019	0,0621	30	0,2014	3	-6,71
	2	0,5491	0,5415	38	-0,0324	-0,0261	27		0,		0,1282	2	-11,05
	3	0,051	0,1121	97	-0,0158	-0,0502	109	0,019	0,0892	128	0,2641	6	-0,42
	4	0,2367	0,2442	24	-0,0569	-0,0759	34	0,019	0,0394	61	0,2707	3	-1,39
	5	0,034	0,1709	126	-0,0141	-0,111	122	0,019	0,2065	138	0,2405	3	-3,93
7	1	0,0858	0,1112	53	0,0202	0,0166	49		0,		0,3583	3	3,04
	2		0,		0,0144	0,0097	67		0,		0,4477	2	0,90
	3	0,3108	2,	195	0,0585	1,0057	215	0,019	0,4575	186	0,1961	9	-6,97
	4	0,0248	0,0543	21	0,0301	0,0235	11		0,		0,1502	2	-12,61
	5	0,2412	0,2593	30	-0,0095	-0,0071	64		0,		0,1433	6	-12,75
9	1		0,		0,0275	0,0199	136	0,019	0,0056	818	0,8349	2	18,26
	2	0,8079	0,8508	13	-0,1456	-0,1282	28		0,		0,3742	1	-0,34
	3	0,1531	0,2028	13	0,0634	0,1276	16	0,019	0,0602	21	0,1524	6	-12,52
	4	0,0747	0,057	19	0,1401	0,2967	10	0,019	0,0857	15	0,1726	4	-9,50
	5	0,5448	0,5634	6	-0,1141	-0,087	19		0,		0,3612	5	2,04
11	1	0,0948	0,0579	59	0,0792	0,234	23	0,019	0,155	31	0,1568	5	-11,23
	2		0,		0,0313	0,3352	274	0,019	0,6351	282	0,1736	3	-8,02
	3	0,3456	0,8681	74	-0,0473	-0,2434	103	0,019	0,1507	110,	0,3555	5	6,12
	4	0,0571	0,0778	16	0,0315	0,0239	12		0,		0,1568	4	-11,83
	5	0,0371	0,0501	43	0,0129	0,0092	41		0,		0,2349	4	-4,84
29	1		0,		-0,0087	-0,0073	96	0,019	0,0111	320	0,1906	4	-8,32
	2	0,9002	2,	200,	0,0221	0,1518	249	0,019	0,1928	313,	0,5309	1	3,13
	3	0,2763	0,2319	29	-0,0668	-0,0844	43	0,019	0,0406	63	0,3841	6	7,82
	4		0,		-0,0121	-0,011	76	0,019	0,0151	189	0,2206	5	-5,69
	5	0,08	0,0841	36	0,0076	0,0059	45		0,		0,1494	5	-12,08
31	1	0,239	0,3328	20	0,0711	0,1367	22	0,019	0,075	33	0,1929	8	-7,49
	2	0,497	0,4549	36	0,0568	0,0433	30		0,		0,2845	4	-3,08
	3	0,078	0,727	937	0,0427	0,6586	969	0,019	0,5246	939	0,3034	4	1,61
	4	0,1086	0,1083	14	0,0898	0,0951	23	0,019	0,0227	62	0,2714	3	-1,35
	5	1,2059	1,2883	16	0,0447	0,0528	39	0,019	0,0315	81	0,1228	4	-14,68
32	1	1,3569	2,	46	-0,074	-0,2039	67	0,019	0,1057	104	0,4782	3	8,85
	2	0,4473	0,4606	457	0,0045	0,0065	371		0,		0,5293	3	3,13
	3	0,4295	0,445	10	-0,0924	-0,1073	21	0,019	0,027	36	0,2615	7	-0,63
	4	0,4807	0,4593	9	-0,0619	-0,0503	20		0,		0,2581	4	-2,87
	5		0,		-0,0149	-0,1354	215	0,019	0,2351	226	0,4717	3	6,31
41	1	0,5142	0,4257	39	0,1078	0,1695	43	0,019	0,0757	67	0,4674	5	8,44
	2	0,9359	0,9932	11	0,1227	0,1131	32		0,		0,3823	1	-0,12
	3	0,3516	0,3505	31	-0,0433	-0,0378	102	0,019	0,0112	314	0,6153	5	18,19

4	0,0666	0,0571	154	0,0195	0,0259	119	0,019	0,0395	241	0,2925	3	0,00	
5	0,161	0,1566	92	0,0028	0,0022	99		0,		0,0983	4	-18,78	
50	1	0,371	0,3853	78	0,0233	0,0202	121	0,019	0,0082	739	0,4057	3	5,89
	2	0,6616	0,7121	76	0,0466	0,0454	83	0,019	0,0126	493	0,3094	3	-2,26
	3	0,0219	0,0209	63	0,0139	0,0148	51	0,019	0,0223	111	0,1493	5	-12,96
	4	0,1737	0,2551	12	0,0625	0,1551	14	0,019	0,0921	19	0,1181	5	-16,33
	5	0,4064	0,4062	17	0,0201	0,0142	32		0,		0,122	5	-15,32

Tabella 15.

Ora abbiamo preso i risultati migliori tra quelli ottenuti nelle Tabelle 14 e 15 ovvero abbiamo scelto le stime per le quali la deviazione standard a posteriori è minore e, a parità di deviazione standard, il coefficiente di Akaike è minore. Molte volte entrambi i valori risultano simili. Se si riscontrano lievissime differenze sono state predilette le stime con maggiore precisione (CV più bassi). Nei casi in cui anche la precisione è risultata identica, abbiamo scelto l'identificazione con un maggior numero di corse. Da ultimo, qualora tutti i valori che abbiamo considerato finora fossero uguali, abbiamo prediletto le stime della seconda tabella.

I risultati sono in Tabella 16.

FFL	S	a _i _iniz	a _i _ott	CV_a _i	b _i _iniz	b _i _ott	CV_b _i	δ _i _iniz	δ _i _ott	CV_δ _i	sd_post	nc	AIC
4	1	0,1764	0,2269	19	-0,1128	-0,2106	21	0,019	0,0748	32	0,2817	5	-0,68
	2	0,8128	0,9111	21	-0,0967	-0,091	52		0,		0,507	1	2,70
	3	0,1616	0,1574	41	-0,015	-0,0136	81	0,019	0,0142	207	0,2284	6	-4,07
	4	0,1671	0,1986	29	-0,0384	-0,0629	39	0,019	0,0512	61	0,2065	3	-6,26
	5	0,4672	0,5261	17	-0,0647	-0,1135	23	0,019	0,0609	36	0,1547	4	-10,98
5	1	0,2511	0,328	17	-0,082	-0,1409	18	0,019	0,0621	30	0,2014	3	-6,71
	2	0,5491	0,5415	38	-0,0324	-0,0261	27		0,		0,1282	2	-11,05
	3	0,051	0,1121	97	-0,0158	-0,0502	109	0,019	0,0892	128	0,2641	6	-0,42
	4	0,2367	0,2442	24	-0,0569	-0,0759	34	0,019	0,0394	61	0,2707	3	-1,39
	5	0,034	0,1709	126	-0,0141	-0,111	122	0,019	0,2065	138	0,2405	3	-3,93
7	1	0,0858	0,1112	53	0,0202	0,0166	49		0,		0,3583	3	3,04
	2		0,		0,0144	0,0097	67		0,		0,4477	2	0,90
	3	0,3108	2,	195	0,0585	1,0057	215	0,019	0,4575	186	0,1961	9	-6,97
	4	0,0248	0,0543	21	0,0301	0,0235	11		0,		0,1502	2	-12,61
	5	0,2412	0,2593	30	-0,0095	-0,0071	64		0,		0,1433	6	-12,75
9	1	0,009	0,0499	163	0,015	0,0247	69		0,		0,818	2	17,90
	2	0,8079	0,8508	13	-0,1456	-0,1282	28		0,		0,3742	1	-0,34
	3	0,1531	0,2028	13	0,0634	0,1276	16	0,019	0,0602	21	0,1524	6	-12,52
	4	0,0747	0,057	19	0,1401	0,2967	10	0,019	0,0857	15	0,1726	4	-9,50
	5	0,5448	0,5634	6	-0,1141	-0,087	19		0,		0,3612	5	2,04

11	1	0,0948	0,0579	59	0,0792	0,234	23	0,019	0,155	31	0,1568	5	-11,23
	2		0,		0,0313	0,3352	274	0,019	0,6351	282	0,1736	3	-8,02
	3	0,3456	0,8681	74	-0,0473	-0,2434	103	0,019	0,1507	110,	0,3555	5	6,12
	4	0,0571	0,0778	16	0,0315	0,0239	12		0,		0,1568	4	-11,83
	5	0,0371	0,0501	43	0,0129	0,0092	41		0,		0,2349	4	-4,84
29	1	0,009	0,0356	187	0,015	-0,0069	56		0,		0,189	4	-8,47
	2	0,009	2,	268	0,015	0,1685	342	0,019	0,2515	425	0,5258	1	3,04
	3	0,2763	0,2319	29	-0,0668	-0,0844	43	0,019	0,0406	63	0,3841	6	7,82
	4		0,		-0,0121	-0,011	76	0,019	0,0151	189	0,2206	5	-5,69
	5	0,08	0,0841	36	0,0076	0,0059	45		0,		0,1494	5	-12,08
31	1	0,239	0,3328	20	0,0711	0,1367	22	0,019	0,075	33	0,1929	8	-7,49
	2	0,497	0,4549	36	0,0568	0,0433	30		0,		0,2845	4	-3,08
	3	0,009	0,8896	1043	0,015	0,8096	1058	0,019	0,6393	1016	0,3032	4	1,59
	4	0,1086	0,1083	14	0,0898	0,0951	23	0,019	0,0227	62	0,2714	3	-1,35
	5	1,2059	1,2883	16	0,0447	0,0528	39	0,019	0,0315	81	0,1228	4	-14,68
32	1	1,3569	2,	46	-0,074	-0,2039	67	0,019	0,1057	104	0,4782	3	8,85
	2	0,4473	0,4606	457	0,0045	0,0065	371		0,		0,5293	3	3,13
	3	0,4295	0,445	10	-0,0924	-0,1073	21	0,019	0,027	36	0,2615	7	-0,63
	4	0,4807	0,4593	9	-0,0619	-0,0503	20		0,		0,2581	4	-2,87
	5		0,		-0,0149	-0,1354	215	0,019	0,2351	226	0,4717	3	6,31
41	1	0,5142	0,4257	39	0,1078	0,1695	43	0,019	0,0757	67	0,4674	5	8,44
	2	0,9359	0,9932	11	0,1227	0,1131	32		0,		0,3823	1	-0,12
	3	0,3516	0,3505	31	-0,0433	-0,0378	102	0,019	0,0112	314	0,6153	5	18,19
	4	0,0666	0,0571	154	0,0195	0,0259	119	0,019	0,0395	241	0,2925	3	0,00
	5	0,161	0,1566	92	0,0028	0,0022	99		0,		0,0983	4	-18,78
50	1	0,371	0,3853	78	0,0233	0,0202	121	0,019	0,0082	739	0,4057	3	5,89
	2	0,6616	0,7121	76	0,0466	0,0454	83	0,019	0,0126	493	0,3094	3	-2,26
	3	0,0219	0,0209	63	0,0139	0,0148	51	0,019	0,0223	111	0,1493	5	-12,96
	4	0,1737	0,2551	12	0,0625	0,1551	14	0,019	0,0921	19	0,1181	5	-16,33
	5	0,4064	0,4062	17	0,0201	0,0142	32		0,		0,122	5	-15,32

Tabella 16

Ora consideriamo i 13 casi in cui uno o più CV sono maggiori di 150%. Per questi ripetiamo l'identificazione dando come valori iniziali le medie dei parametri ottenute considerando solo i casi "buoni", ovvero quelli con CV accettabili.

Non si ottengono miglioramenti.

Allora come passo successivo fissiamo il parametro di degradazione alla media calcolata (sempre considerando i casi "buoni").

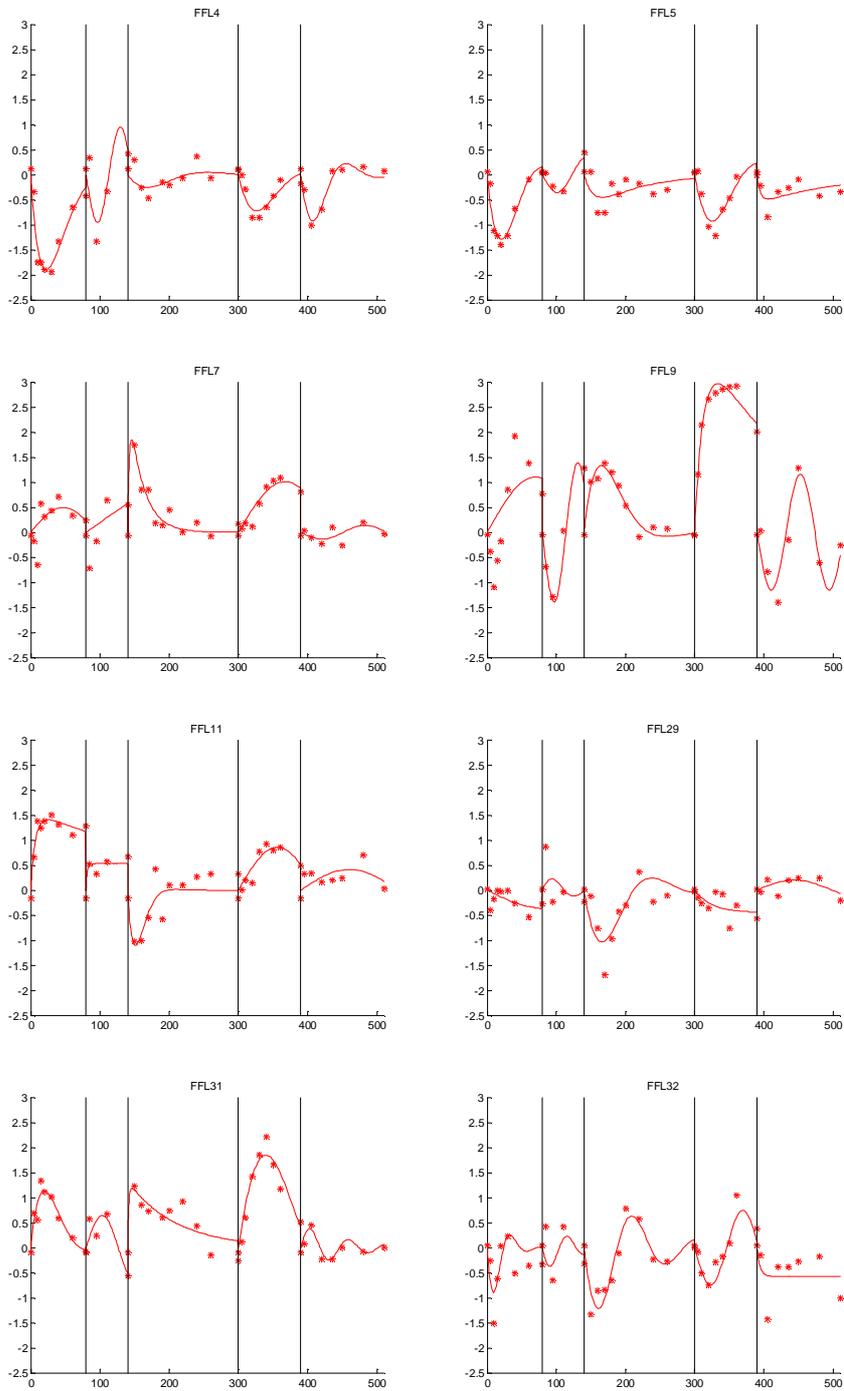
Alcuni casi migliorano e decidiamo di fermarci qui (Tabella 17).

FFL	S	a_{i_iniz}	a_{i_ott}	CV_{a_i}	b_{i_iniz}	b_{i_ott}	CV_{b_i}	δ_{i_iniz}	δ_{i_ott}	CV_{δ_i}	sd_post	nc	AIC
4	1	0,1764	0,2269	19	-0,1128	-0,2106	21	0,019	0,0748	32	0,2817	5	-0,68
	2	0,8128	0,9111	21	-0,0967	-0,091	52		0,		0,507	1	2,70
	3	0,3817	0,1766	69	-0,0061	-0,0195	60		0,0413		0,2257	6	-4,98

	4	0,1671	0,1986	29	-0,0384	-0,0629	39	0,019	0,0512	61	0,2065	3	-6,26
	5	0,4672	0,5261	17	-0,0647	-0,1135	23	0,019	0,0609	36	0,1547	4	-10,98
5	1	0,2511	0,328	17	-0,082	-0,1409	18	0,019	0,0621	30	0,2014	3	-6,71
	2	0,5491	0,5415	38	-0,0324	-0,0261	27		0,		0,1282	2	-11,05
	3	0,051	0,1121	97	-0,0158	-0,0502	109	0,019	0,0892	128	0,2641	6	-0,42
	4	0,2367	0,2442	24	-0,0569	-0,0759	34	0,019	0,0394	61	0,2707	3	-1,39
	5	0,034	0,1709	126	-0,0141	-0,111	122	0,019	0,2065	138	0,2405	3	-3,93
7	1	0,0858	0,1112	53	0,0202	0,0166	49		0,		0,3583	3	3,04
	2		0,		0,0144	0,0097	67		0,		0,4477	2	0,90
	3	0,3108	2,	195	0,0585	1,0057	215	0,019	0,4575	186	0,1961	9	-6,97
	4	0,0248	0,0543	21	0,0301	0,0235	11		0,		0,1502	2	-12,61
	5	0,2412	0,2593	30	-0,0095	-0,0071	64		0,		0,1433	6	-12,75
9	1	0,009	0,0499	163	0,015	0,0247	69		0,		0,818	2	17,90
	2	0,8079	0,8508	13	-0,1456	-0,1282	28		0,		0,3742	1	-0,34
	3	0,1531	0,2028	13	0,0634	0,1276	16	0,019	0,0602	21	0,1524	6	-12,52
	4	0,0747	0,057	19	0,1401	0,2967	10	0,019	0,0857	15	0,1726	4	-9,50
	5	0,5448	0,5634	6	-0,1141	-0,087	19		0,		0,3612	5	2,04
11	1	0,0948	0,0579	59	0,0792	0,234	23	0,019	0,155	31	0,1568	5	-11,23
	2		0,		0,0313	0,3352	274	0,019	0,6351	282	0,1736	3	-8,02
	3	0,3456	0,8681	74	-0,0473	-0,2434	103	0,019	0,1507	110,	0,3555	5	6,12
	4	0,0571	0,0778	16	0,0315	0,0239	12		0,		0,1568	4	-11,83
	5	0,0371	0,0501	43	0,0129	0,0092	41		0,		0,2349	4	-4,84
29	1	0,009	0,0356	187	0,015	-0,0069	56		0,		0,189	4	-8,47
	2	0,3817	1,0403	138	-0,0061	0,0304	205		0,0413		0,515	1	2,86
	3	0,2763	0,2319	29	-0,0668	-0,0844	43	0,019	0,0406	63	0,3841	6	7,82
	4		0,		-0,0061	-0,0186	23		0,0413		0,2113	5	-7,27
	5	0,08	0,0841	36	0,0076	0,0059	45		0,		0,1494	5	-12,08
31	1	0,239	0,3328	20	0,0711	0,1367	22	0,019	0,075	33	0,1929	8	-7,49
	2	0,497	0,4549	36	0,0568	0,0433	30		0,		0,2845	4	-3,08
	3	0,009	0,8896	1043	0,015	0,8096	1058	0,019	0,6393	1016	0,3032	4	1,59
	4	0,1086	0,1083	14	0,0898	0,0951	23	0,019	0,0227	62	0,2714	3	-1,35
	5	1,2059	1,2883	16	0,0447	0,0528	39	0,019	0,0315	81	0,1228	4	-14,68
32	1	1,3569	2,	46	-0,074	-0,2039	67	0,019	0,1057	104	0,4782	3	8,85
	2	0,3817	1,7014	115	-0,0061	-0,0613	103		0,0413		0,4669	3	1,88
	3	0,4295	0,445	10	-0,0924	-0,1073	21	0,019	0,027	36	0,2615	7	-0,63
	4	0,4807	0,4593	9	-0,0619	-0,0503	20		0,		0,2581	4	-2,87
	5		0,		-0,0149	-0,1354	215	0,019	0,2351	226	0,4717	3	6,31
41	1	0,5142	0,4257	39	0,1078	0,1695	43	0,019	0,0757	67	0,4674	5	8,44
	2	0,9359	0,9932	11	0,1227	0,1131	32		0,		0,3823	1	-0,12
	3	0,3817	0,3496	57	-0,0061	-0,0562	65		0,0413		0,5857	5	16,40
	4	0,3817	0,0561	124	-0,0061	0,0265	39		0,0413		0,2708	3	-2,00
	5	0,161	0,1566	92	0,0028	0,0022	99		0,		0,0983	4	-18,78
50	1	0,3817	0,3563	107	-0,0061	0,0278	78		0,0413		0,3833	2	4,25
	2	0,3817	0,6211	72	-0,0061	0,0559	54		0,0413		0,268	2	-3,67
	3	0,0219	0,0209	63	0,0139	0,0148	51	0,019	0,0223	111	0,1493	5	-12,96
	4	0,1737	0,2551	12	0,0625	0,1551	14	0,019	0,0921	19	0,1181	5	-16,33
	5	0,4064	0,4062	17	0,0201	0,0142	32		0,		0,122	5	-15,32

Tabella 17

I plot corrispondenti ai fit della predizione del modello contro i dati sono riportati di seguito (Figura 4.9).



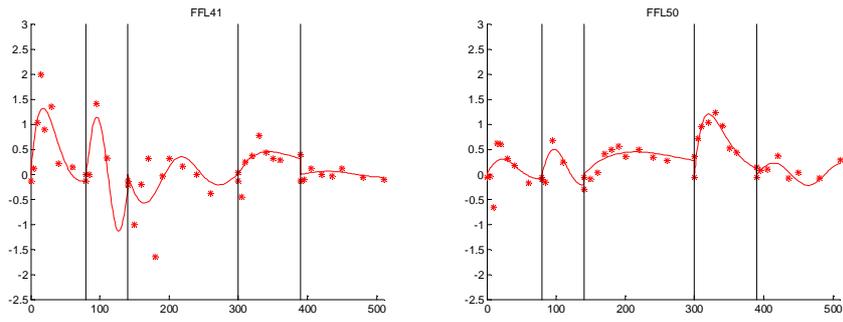


Figura 4.9

Capitolo 5

Risultati

5.1. Risultati

Dal lavoro svolto nel paragrafo 4.1.1 abbiamo ottenuto una tabella finale che riporta i risultati migliori in termini di fit, derivanti dalle varie identificazioni fatte partendo dal modello in cui il parametro di degradazione k_{03} era stato fissato. Allo stesso modo nel paragrafo 4.1.2 si è ottenuta una tabella finale con i risultati migliori. A questo punto abbiamo confrontato i coefficienti di Akaike delle due tabelle e abbiamo scelto i risultati della stima parametrica che conducono ad un minore coefficiente di Akaike.

Lo stesso confronto è stato fatto sulle identificazioni relative ai dati dei FFL 22 e FFL 25.

Nella tabella 1 sono riportate le stime ottime, la precisione delle stime, l'SD stimata a posteriori, il numero di corse e l'indice di Akaike ottenuti per i 12 FFL analizzati (per ogni FFL sono state considerate 5 serie temporali in risposta allo stress). Nella seconda colonna della tabella abbiamo indicato con una "x" le righe in cui i parametri hanno CV elevati o pari a zero e, in fondo alla tabella, abbiamo riportato i valori medi per ogni colonna, escludendo questi casi.

Sono state poi calcolate anche le medie ottenute considerando tutti i valori.

FFL	S	k_{31_ott}	CV_k_{31}	k_{32_ott}	CV_k_{32}	k_{03_ott}	CV_k_{03}	sd_post	nc	AIC	
4	1	0,2504	81	0,1543	90	0,1566	37	0,2376	4	-3,75	X
	2	0,2092	101	0,3556	83	0,1515	120	0,4948	3	2,43	≈
	3	0,0013		0,0036	2	0,019		0,2541	4	-3,43	
	4	0,14	26	0,1978	16	0,3163	2	0,1489	5	-12,16	X
	5	0,0351	73	0,2448	14	0,0624	22	0,1362	4	-13,03	X

5	1	-0,0741	86	0,3375	25	0,1532	23	0,1611	5	-10,73	X
	2	0,0704	10	0,0323	16	0,		0,0573	2	-19,10	X
	x 3	0,0917	3	0,1513	0	0,2026	4	0,2288	6	-3,57	X
	4	0,2781	23	0,3274	60	0,4255	54	0,2213	2	-5,02	X
	5	-0,0261	66	0,2154	9	0,019		0,2532	4	-3,64	≈
7	1	0,0013		-0,0292	30	0,019		0,3204	4	0,23	X
	2	0,0616	128	-0,0447	94	0,019		0,4813	2	2,18	
	3	0,2016	2	-0,0262	13	0,019		0,3735	7	6,51	
	4	-0,0477	29	-0,0274	17	0,019		0,105	3	-19,05	X
	5	0,0762	18	-0,1316	24	0,2091	47	0,1281	6	-14,00	X
9	1	-0,3639	8	0,3502	15	0,0592	30	0,416	6	6,34	X
	2	0,5921	30	0,4722	44	0,1152	40	0,251	3	-4,36	X
	3	0,1937	5	0,1538	15	0,019		0,6126	2	17,39	
	x 4	0,1804	13	-0,4833	10	0,001	738	0,3815	3	4,78	
	5	0,0953	94	-0,1911	116	0,0818	111	0,5878	4	10,37	
11	1	-0,3257	44	-0,2867	22	0,019		0,6786	3	14,54	
	2	0,1413	34	0,0636	72	0,		0,2643	3	-3,82	
	3	-0,1764	30	-0,0513	16	0,019		0,5399	4	14,61	
	4	0,1337	21	0,0136	145	0,019		0,2115	4	-6,45	
	5	0,0532	48	0,2022	26	0,0892	33	0,1845	5	-8,16	X
29	1	0,0013		-0,0161	20	0,019		0,1774	4	-10,41	≈
	x 2	-0,0019		-0,014	215	0,019		0,4544	1	1,04	X
	x 3	-0,5895	3	-0,1513	0	0,2777	0	0,4367	5	10,65	
	4	-0,0142	125	-0,0135	101	0,019		0,2448	5	-3,82	
	5	-0,0019		0,012	58	0,019		0,1865	5	-9,30	
31	x 1	-0,1028	291	0,6639	15	0,5257	10	0,3736	3	4,41	
	2	0,0975	82	-0,0704	39	0,		0,3332	4	-1,50	
	3	0,1044	61	0,0061	53	0,019		0,5407	4	12,50	
	4	0,5621	23	-0,1023	14	0,019		0,6105	4	12,63	
	5	0,0013		-0,0095	71	0,019		0,2132	5	-7,16	
32	1	0,0013		-0,0191	87	0,019		0,5859	4	11,09	
	x 2	0,0013		-0,0106	316	0,019		0,4585	3	1,13	≈
	x 3	-0,352	84	-0,3989	10	0,4608	0	0,5638	3	16,27	
	4	-0,208	11	0,0727	32	0,019		0,4013	3	5,08	
	5	0,0013		-0,0963	32	0,019		0,4662	5	5,36	X
41	1	0,2185	37	0,373	17	0		0,4681	5	7,85	X
	x 2	-1,1441	133	-1,7543	530	0,7948	71	0,5903	3	4,19	
	3	0,0013		0,0074	92	0,019		0,6269	5	17,05	
	4	-0,0105	35	0,0573	45	0,019		0,2718	3	-1,94	≈
	5	0,		0,033	45	0,		0,0857	5	-21,74	X
50	1	-0,0528	58	0,0645	47	0,019		0,3679	3	3,52	≈
	x 2	-0,5389	172	0,2536	188	0,2606	205	0,2996	2	-2,59	
	x 3	0,0255	74	-0,004	231	0,0079	90	0,1155	3	-18,61	X
	4	-0,2431	32	0,4009	13	0,253	7	0,2122	5	-5,78	
	5	0,0508	113	0,047	100	0,1719	104	0,1462	7	-11,89	
22	1	-0,4809	58	-0,3873	35	0,019		1,2819	3	25,98	
	2	0,1315	42	0,3236	13	0,019		0,2009	2	-6,56	X
	3	-0,0885	51	-0,0439	7	0,019		0,6002	4	16,94	
	4	0,0749	4	-0,0066	8	0,019		0,275	4	-1,72	
	5	0,0181	63	0,0331	64	0,019		0,1896	4	-8,27	≈
25	1	-0,6388	24	-0,5452	12	0,019		1,1122	3	23,43	
	2	1,4394	4	-0,234	38	0,1866	24	0,2127	2	-6,01	X
	3	0,1901	16	0,007	66	0,019		0,3413	5	4,52	
	4	0,4505	71	-0,2051	38	0,0773	73	1,0552	4	23,09	
	5	0,3107	27	0,1855	49	0,1084	34	0,3158	5	0,43	X

Medie dei fit buoni

0,0687	46	0,0442	43	0,0634	48	0,3729	4	0,43
--------	----	--------	----	--------	----	--------	---	------

Medie di tutti i valori

0,0168	54	0,0077	61	0,0956	78	0,3758	4	0,65
--------	----	--------	----	--------	----	--------	---	------

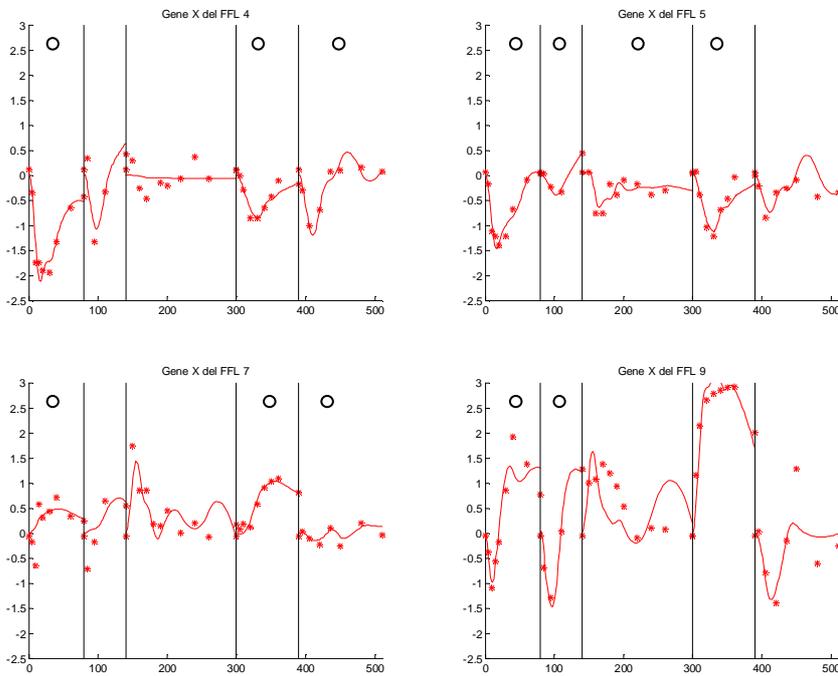
Medie fit buoni fatte per serie

S	k_{31_ott}	CV_ k_{31}	k_{32_ott}	CV_ k_{32}	k_{03_ott}	CV_ k_{03}	sd_post	nc	AIC
1	-0,133	50	-0,0004	36	0,0456	30	0,5279	4	6,19
2	0,3429	54	0,1123	50	0,0614	61	0,2869	3	-4,59
3	0,0534	28	0,0071	33	0,019		0,4862	4	10,76
4	0,1014	36	0,065	44	0,1096	34	0,3416	4	-1,38
5	0,0512	63	0,0454	51	0,0682	59	0,2411	5	-6,75

Medie di tutti i valori fatte per serie

S	k_{31_ott}	CV_ k_{31}	k_{32_ott}	CV_ k_{32}	k_{03_ott}	CV_ k_{03}	sd_post	nc	AIC
1	-0,1305	76	0,055	35	0,0856	25	0,5151	4	6,04
2	0,0883	74	-0,0522	137	0,1321	92	0,3415	3	-2,75
3	-0,0331	33	-0,0288	42	0,0917	24	0,4362	4	7,57
4	0,108	34	0,0193	42	0,1005	175	0,3449	4	-0,86
5	0,0512	63	0,0454	51	0,0682	59	0,2411	5	-6,75

Tabella 1



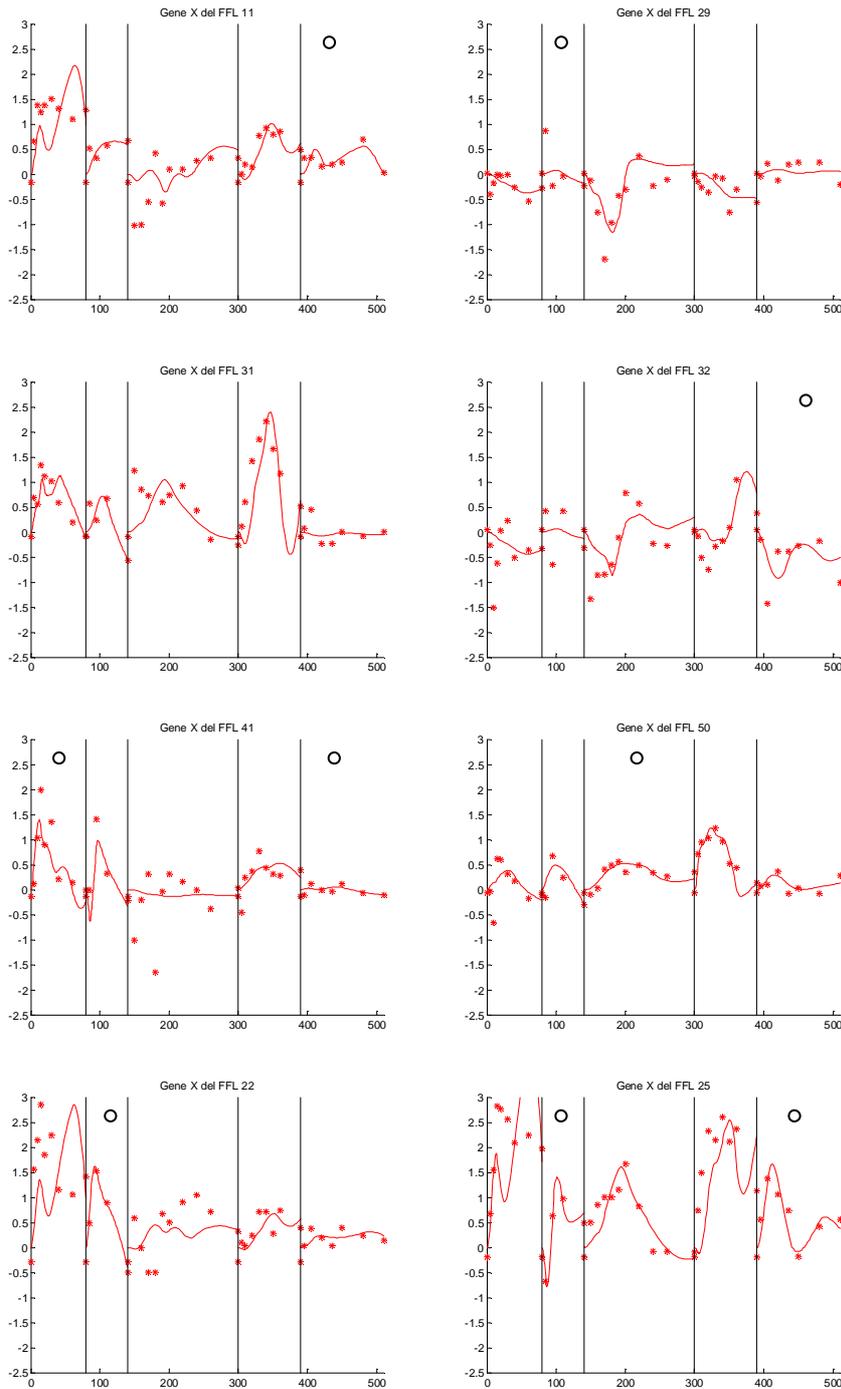


Figura 5.1

Nella Tabella 1 si ha dunque una panoramica delle 60 identificazioni del modello proposto in questo lavoro di tesi. La procedura descritta nei paragrafi 4.1.1 e 4.1.2 ha consentito di individuare dei fit via via migliori e ci ha permesso di capire quali dei parametri si possono

stimare, quali invece si devono imporre e quali altri risultano praticamente nulli, nei vari casi studiati.

Allo stesso modo si è ottenuta una tabella riassuntiva delle 60 identificazioni del modello proposto da Altafini (Tabella 2).

FFL	S	a _{ott}	CV _{a_i}	b _{ott}	CV _{b_i}	δ _{ott}	CV _{δ_i}	sd _{post}	nc	AIC	
4	1	0,2269	19	-0,2106	21	0,0748	32	0,2817	5	-0,68	≈ X
	2	0,9111	21	-0,091	52	0,		0,507	1	2,70	
	3	0,1616	41	-0,015	47	0,019		0,2145	7	-5,99	
	4	0,1986	29	-0,0629	39	0,0512	61	0,2065	3	-6,26	
	5	0,5261	17	-0,1135	23	0,0609	36	0,1547	4	-10,98	
5	1	0,328	17	-0,1409	18	0,0621	30	0,2014	3	-6,71	≈
	2	0,5415	38	-0,0261	27	0,		0,1282	2	-11,05	
	3	0,1121	97	-0,0502	109	0,0892	128	0,2641	6	-0,42	
	4	0,2442	24	-0,0759	34	0,0394	61	0,2707	3	-1,39	
	5	0,1709	126	-0,111	122	0,2065	138	0,2405	3	-3,93	
7	1	0,1112	53	0,0166	49	0,		0,3583	3	3,04	X X
	2	0,		0,0097	67	0,		0,4477	2	0,90	
	3	2,	195	1,0057	215	0,4575	186	0,1961	9	-6,97	
	4	0,0543	21	0,0235	11	0,		0,1502	2	-12,61	
	5	0,2593	30	-0,0071	64	0,		0,1433	6	-12,75	
9	1	0		0,0275	41	0,019		0,7902	4	16,48	X X X
	2	0,8508	13	-0,1282	28	0,		0,3742	1	-0,34	
	3	0,2028	13	0,1276	16	0,0602	21	0,1524	6	-12,52	
	4	0,057	19	0,2967	10	0,0857	15	0,1726	4	-9,50	
	5	0,5634	6	-0,087	19	0,		0,3612	5	2,04	
11	1	0,0579	59	0,234	23	0,155	31	0,1568	5	-11,23	X X X X
	2	0,		0,3352	274	0,6351	282	0,1736	3	-8,02	
	3	0,8681	74	-0,2434	103	0,1507	110,	0,3555	5	6,12	
	4	0,0778	16	0,0239	12	0,		0,1568	4	-11,83	
	5	0,0501	43	0,0092	41	0,		0,2349	4	-4,84	
29	1	0		-0,0087	29	0,019		0,179	6	-10,26	≈ X X X
	2	1,0403	138	0,0304	205	0,0413		0,515	1	2,86	
	3	0,2763	17	-0,0668	24	0,019		0,3962	4	7,80	
	4	0		-0,0121	22	0,019		0,2066	6	-7,67	
	5	0,0841	36	0,0059	45	0,		0,1494	5	-12,08	
31	1	0,3328	20	0,1367	22	0,075	33	0,1929	8	-7,49	X X X X X
	2	0,4549	36	0,0433	30	0,		0,2845	4	-3,08	
	3	0,8896	1043	0,8096	1058	0,6393	1016	0,3032	4	1,59	
	4	0,1086	12	0,0898	8	0,019		0,2531	5	-3,22	
	5	1,2059	8	0,0447	23	0,019		0,1143	6	-16,37	
32	1	2,	46	-0,2039	67	0,1057	104	0,4782	3	8,85	X ≈ X X
	2	1,7014	115	-0,0613	103	0,0413		0,4669	3	1,88	
	3	0,4295	9	-0,0924	13	0,019		0,2605	8	-1,43	
	4	0,4593	9	-0,0503	20	0,		0,2581	4	-2,87	
	5	0,		-0,1354	215	0,2351	226	0,4717	3	6,31	
41	1	0,4257	39	0,1695	43	0,0757	67	0,4674	5	8,44	X X ≈
	2	0,9932	11	0,1131	32	0,		0,3823	1	-0,12	
	3	0,3516	35	-0,0433	59	0,019		0,5809	7	16,22	
	4	0,0561	124	0,0265	39	0,0413		0,2708	3	-2,00	
	5	0,1566	92	0,0022	99	0,		0,0983	4	-18,78	

50	1	0,371	81	0,0233	70	0,019		0,3764	6	3,93	≈
	2	0,6616	60	0,0466	56	0,019		0,2535	3	-4,23	X
	3	0,0219	43	0,0139	17	0,019		0,1411	7	-14,91	
	4	0,2551	12	0,1551	14	0,0921	19	0,1181	5	-16,33	X
	5	0,4062	17	0,0142	32	0,		0,122	5	-15,32	X

22	1	0,2809	42	0,5484	42	0,21	50	0,4476	5	7,66	X
	2	0,5649	26	0,1587	27	0,0451	46	0,2694	3	-3,65	
	3	0,		0,0124	34	0,019		0,4736	8	10,89	X
	4	0,058	56	0,0237	24	0,019		0,2172	5	-5,97	X
	5	0,0412	84	0,0101	46	0,019		0,1854	5	-8,63	≈

25	1	0,0745	45	0,2917	29	0,0958	44	0,398	3	5,54	X
	2	0,1752	55	0,0371	58	0,		0,5155	2	2,87	
	3	0,0864	14	0,0555	11	0,019		0,2669	7	-0,89	X
	4	0,0862	15	0,1806	15	0,0502	25	0,2318	6	-4,19	X
	5	0,3626	58	0,2108	67	0,1257	81	0,3948	3	4,00	

Medie dei fit buoni

0,3465	40	0,0233	40	0,0418	57	0,2868	4	-3,15
--------	----	--------	----	--------	----	--------	---	-------

Medie di tutti valori

0,3831	64	0,0554	69	0,0718	118	0,2906	4	-2,96
--------	----	--------	----	--------	-----	--------	---	-------

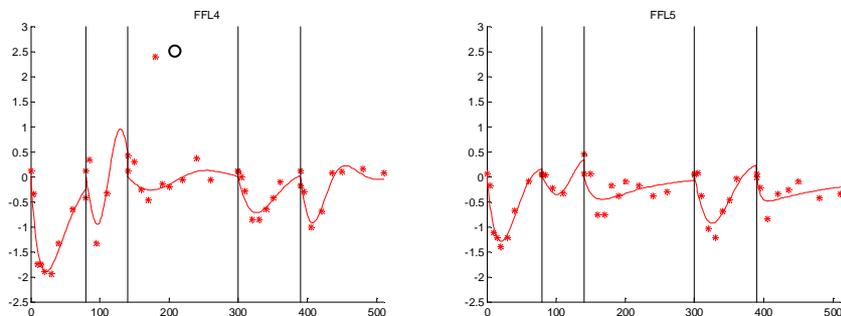
Medie dei fit buoni fatte per serie

S	ai_ott	CV_ai	bi_ott	CV_bi	δi_ott	CV_δi	sd_post	nc	AIC
1	0,3507	42	0,0736	38	0,0759	49	0,3607	5	1,46
2	0,6855	42	0,0102	48	0,0105	46	0,3629	2	-1,41
3	0,251	38	-0,0302	43	0,0433	86	0,3106	7	0,49
4	0,1379	31	0,0516	21	0,0347	36	0,2094	4	-6,99
5	0,3479	47	-0,0019	53	0,0392	85	0,1999	5	-8,88

Medie di tutti i valori fatte per serie

S	ai_ott	CV_ai	bi_ott	CV_bi	δi_ott	CV_δi	sd_post	nc	AIC
1	0,3507	42	0,0736	38	0,0759	49	0,3607	5	1,46
2	0,6579	51	0,039	80	0,0651	164	0,3598	2	-1,61
3	0,45	144	0,1261	142	0,1275	292	0,3004	7	-0,04
4	0,1379	31	0,0516	21	0,0347	36	0,2094	4	-6,99
5	0,3189	47	-0,0131	66	0,0555	120	0,2226	4	-7,61

Tabella 2



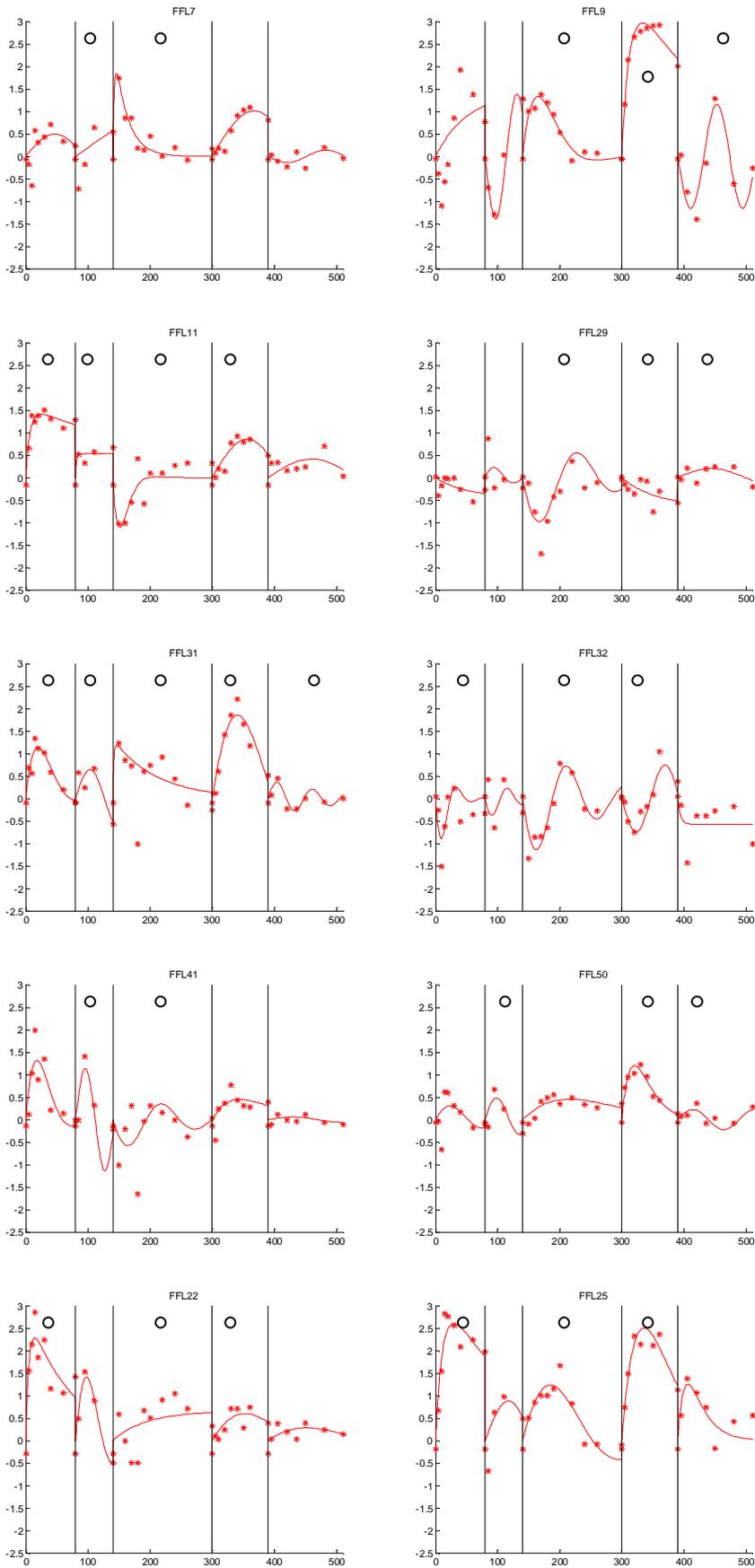


Figura 5.2

Confrontiamo i coefficienti di Akaike delle due tabelle. Nell'ultima colonna di ogni tabella sono stati segnati con una "X" i casi in cui il coefficiente di Akaike è minore in un modello piuttosto che nell'altro (nei plot le serie corrispondenti sono contrassegnate da un pallino). Con il simbolo "≈" abbiamo invece indicato i casi in cui i valori sono simili per entrambi i modelli.

La situazione è la seguente: in 32 casi su 60 è minore il coefficiente di Akaike nel modello con feedback integrativo proposto da Altafini, in 21 casi su 60 è minore quello del modello basato su feed-forward loop, proposto nel presente lavoro di tesi, e nei restanti 7 casi i valori sono molto simili.

In media, però, il coefficiente di Akaike risulta minore nel modello di Altafini (-2,96 vs 0,65).

Per quanto riguarda la deviazione standard a posteriori i valori medi sono:

$SD_{post} = 0,2906$ per il modello "feedback-integrativo",

$SD_{post} = 0,3758$ per il modello "feed-forward loop".

Sono entrambi inferiori a 0,4 e ciò rappresenta un risultato accettabile. Anche le precisioni medie delle stime in termini di CV sono accettabili: 54%, 61% e 78% nel caso del nostro modello e 64%, 69% e 118% nel caso del modello Altafini. L'unico valore meno soddisfacente è il CV medio pari a 118% del parametro δ_i .

Spesso vediamo che il parametro di degradazione viene fissato a zero. Questo molto probabilmente è spiegabile col fatto che i tempi di degradazione possono essere superiori alla finestra temporale che si prende in considerazione. Notiamo inoltre che, quando è diverso da zero, spesso non si riesce a stimarlo. Ci sono molte situazioni, in entrambi i modelli, in cui abbiamo dovuto fissarlo. Un'altra considerazione importante relativa al parametro δ_i o equivalentemente k_{03} , è che il valor medio, calcolato sui 60 casi studiati, è superiore al valore medio trovato da Altafini. Infatti risulta: $k_{03_medio} \cong 0,0956$ e $\delta_{i_medio} \cong 0,0718$ mentre nel lavoro di Altafini [3] $\delta_{i_medio} \cong 0,019$.

Nelle identificazioni del modello di Altafini spesso il parametro a_i , che rappresenta la forza della regolazione della proteina sulla produzione del gene target, assume valore nullo. Quindi la retroazione scompare e questo significa che, in base alla struttura del modello, l'unica azione in grado di far riassorbire il picco transitorio è la degradazione.

Notiamo che il modello "feed-forward" riesce a descrivere anche risposte allo stress con andamento diverso rispetto a quello caratterizzato da un picco iniziale seguito dal ritorno al livello basale. Ad esempio nella serie 4 del FFL 7 l'mRNA rimane al livello di concentrazione basale nei primi 10 minuti circa e poi si ha una sovraespressione. Nella serie 1 del FFL 9,

invece, il gene viene sottoespresso e poi sovraespresso e, anche in questo caso, il modello “feedback-integrativo” riesce a descrivere meglio i dati rispetto al modello di Altafini.

Capitolo 6

Conclusioni

6.1. Conclusioni

I processi cellulari coinvolgono milioni di molecole che svolgono un ruolo coerente al fine di scambiare materia, energia e informazione con l'ambiente. Questi processi vengono regolati dai geni, la cui espressione è a loro volta regolata da una rete di interazioni tra altri geni, proteine e molecole, detta rete di regolazione genica. Data la complessità della rete non è difficile apprezzare la difficoltà di una loro modellizzazione matematica.

La nostra attenzione si è concentrata sul tipico profilo di espressione in risposta allo stress (ESR), comune a un gran numero di geni appartenenti al genoma del *S. Cerevisiae* e caratterizzato da un picco di espressione e un successivo ritorno al livello basale. Nel presente lavoro è stato definito un sistema di equazioni differenziali che descrivono il modello di regolazione proposto per descrivere la dinamica di uno dei motivi-base della regolazione, il feed-forward loop.

Assumendo come ipotesi che la risposta allo stress possa essere spiegata da un meccanismo di regolazione come quello descritto dal feed-forward loop, è stata eseguita l'identificazione parametrica del modello proposto, sfruttando i dati di espressione relativi ad un set di geni con il tipico profilo. I risultati della stima parametrica hanno dimostrato che, in media, il modello predice i dati in modo soddisfacente. Usando lo stesso data set è stato identificato anche il modello "feedback integrativo" proposto dal gruppo di Altafini [3] appositamente per spiegare la risposta allo stress. Dal lavoro di tesi risulta che entrambi i modelli sono in grado di descrivere la risposta allo stress in maniera comparabile. Il modello "feed-forward loop" proposto permette, però, di tenere in considerazione la rete trascrizionale che sta alla base della regolazione genica e di descrivere anche pattern di risposta allo stress che si discostano da quello tipico. Interpretare i dati di

espressione genica tenendo conto della complessa rete di regolazione è sicuramente un approccio molto arduo ma che può condurre alla costruzione di modelli più aderenti alla realtà biologica.

Un possibile sviluppo di questo lavoro potrebbe essere dato dall'identificazione degli stessi due modelli usando un maggior numero di geni e sarebbe anche interessante vedere come i due modelli descrivono profili di espressione diversi da quello considerato in questa tesi.

Bibliografia

- [1] Genomic expression programs in the response of yeast cells to environmental changes. **A. P. Gasch et al.** 2000, *Mol. Biol. Cell*, Vol. 11, p. 4241-4257.
- [2] Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. **G. Chechik et al.** *Nature Biotechnology*, 2008, Vol. 26, no. 11, p. 1251-1259.
- [3] Adaptation as a genome-wide autoregulatory principle in the stress response of yeast. **G. De Palo et al.** 2010 (sottomesso a *IET Systems Biology*)
- [4] **C. Cobelli e R. Bonadonna.** *Bioingegneria dei sistemi metabolici*. Bologna: Pàtron editore, 1998
- [5] Network motifs in the transcriptional regulation of *Escherichia Coli*. **S. S. Shen-Orr et al.** 2002, *Nature genetics*, Vol. 31, p. 64-68.
- [6] Network motifs: theory and experimental approaches. **U. Alon.** 2007, *Nature*, Vol. 8, p. 450-461.
- [7] Structure and function of the feed-forward loop network motif. **S. Mangan and U. Alon.** 2003, *Proc Natl Acad Sci USA*, Vol. 100, no. 21, p. 11980-11985.
- [8] Network Motifs: Simple Building Blocks of Complex Networks. **R. Milo et al.** 2002, *Science*, Vol. 298, p. 824-827.
- [9] Transcriptional regulatory networks in *Saccharomyces cerevisiae*. **T. I. Lee et al.** 2002, *Science*, Vol. 298, p. 799-804.
- [10] **R. Bellazzi et al.** *Genomica e proteomica computazionale*. Bologna: Pàtron editore, 2007.
- [11] **A. Zhang.** *Advanced analysis of gene expression microarray data*. World Scientific, 2006.

Ringrazio la mia famiglia ed Andrea.