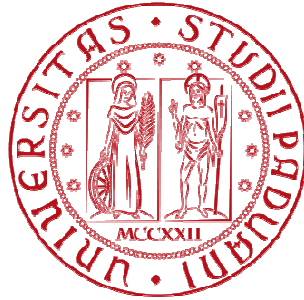


**UNIVERSITÀ DEGLI STUDI DI PADOVA**  
FACOLTÀ DI INGEGNERIA  
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE (DEI)



**Tesi di Laurea Magistrale in Bioingegneria**

**MODELLI DINAMICI DELLA FOSFORILAZIONE NEL  
SIGNALING DELL'INSULINA**

*Relatore:* Prof.ssa Gianna Maria Toffolo

*Correlatori:* Prof.ssa Barbara Di Camillo

Ing. Federica Eduati

*Laureando:* RONNYE PAJARO

ANNO ACCADEMICO 2011-2012



## Sommario

La rete di signaling dell'insulina, localizzata all'interno della cellula, è costituita da molecole biologiche, tipicamente di natura proteica, che interagiscono tra loro attraverso meccanismi più o meno complessi, come ad esempio meccanismi di feed-forward o di feedback. La reazione chimica fondamentale che interessa le proteine di questa rete è la fosforilazione, che consiste nell'aggiunta di un gruppo fosfato a un residuo amminoacidico della catena polipeptidica. Per la maggior parte delle proteine, la fosforilazione della catena ne determina l'attivazione, ossia la trasforma in una chinasi, cioè in un enzima in grado di favorire, a sua volta, la fosforilazione di un'altra proteina. Nella minoranza dei casi, invece, la fosforilazione agisce in maniera opposta, ossia determina l'inattivazione della proteina. La rete di signaling dell'insulina è coinvolta in eventi biologici fondamentali, quali la sintesi proteica, la proliferazione e la differenziazione cellulare, l'utilizzazione del glucosio e la sintesi del glicogeno. L'insulina è un ormone presente all'esterno della cellula e il segnale da essa trasportato è trasferito all'interno attraverso il corrispondente recettore di membrana (IR).

Il presente lavoro di tesi si pone l'obiettivo di realizzare dei modelli matematici dei moduli di signaling che compongono la rete in questione, dove per modulo di signaling s'intende un sotto-sistema elementare della rete, costituito da certa proteina, oggetto della fosforilazione, e dalla corrispondente proteina che agisce da controllore. I dati di riferimento sono stati acquisiti presso la Mayo Clinic (Progetto CARIPARO) su culture cellulari di mioblasti umani di muscolo scheletrico, sfruttando la tecnica di misura western blot. Le misure rappresentano un'intensità che risulta proporzionale alla quantità (massa in  $[\mu\text{g}]$ ) di proteina e sono state ottenute in corrispondenza di 3 differenti stimoli: insulina, leucina e leucina+insulina. I parametri univocamente identificabili a priori dei modelli compartimentali scelti vengono quindi opportunamente identificati attraverso il metodo dei minimi quadrati (non lineari) pesati. L'obiettivo futuro sarà quello di assemblare la rete globale a partire dai singoli moduli.



# Indice

## **INTRODUZIONE**

### **1 INSULINA E INSULIN SIGNALING PATHWAY**

1.1 Struttura della molecola di insulina

1.2 Produzione

1.3 Secrezione

1.4 Insulin signaling pathway

1.4.1 Introduzione

1.4.2 Componenti del pathway

1.4.3 Terminazione del pathway

### **2 DATI ACQUISITI**

2.1 Configurazione sperimentale

2.2 Misure

2.2.1 Tecnica utilizzata: western blot

2.2.2 Proteine misurate

2.2.3 Repliche disponibili

2.2.4 Grafici delle serie temporali

2.2.5 Valutazione qualitativa dei dati

### **3 PRE-ELABORAZIONE DEI DATI**

3.1 Modello dell'errore tecnico e biologico

3.2 Stima del pattern attraverso l'operazione di media

3.3 Stima del pattern attraverso un algoritmo Expectation-Maximization

### **4 MODELLI**

4.1 Modelli di un singolo modulo di signaling

4.2 Equazioni del modello con controllo sulla fosforilazione

4.3 Equazioni del modello con controllo sulla defosforilazione

4.4 Equazione ingresso/uscita del modello

### **5 IDENTIFICAZIONE DEI PARAMETRI**

5.1 Descrizione teorica: stima ai minimi quadrati pesati

5.2 Implementazione

### **6 RISULTATI**

6.1 Dati medi

6.2 Pattern Expectation-Maximization

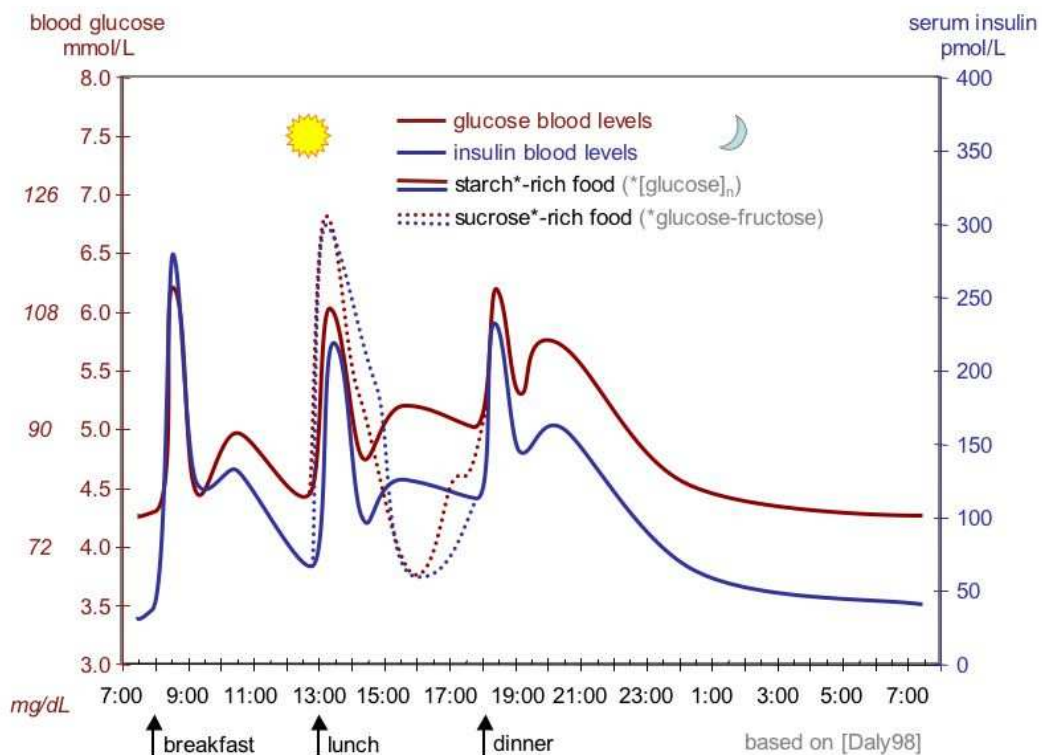
6.3 Discussione dei risultati

7 **CONCLUSIONI**

**BIBLIOGRAFIA**

# INTRODUZIONE

L'insulina è un ormone essenziale per una normale crescita, un normale sviluppo e una normale omeostasi del metabolismo del glucosio, dei grassi e delle proteine. In particolare, l'omeostasi del glucosio dipende dal bilancio tra la produzione del glucosio epatico e l'utilizzazione del glucosio da parte dei maggiori tessuti insulino-dipendenti, come il fegato, il tessuto muscolare e il tessuto adiposo, e da parte dei tessuti insulino-indipendenti, come il cervello e i reni. Questo bilancio è strettamente regolato dall'insulina e dagli altri ormoni pancreatici, come ad esempio il glucagone. Perciò, in individui normali, la risposta ad un incremento dei livelli di glucosio plasmatico consiste in un incremento della secrezione di insulina dalle  $\beta$ -cellule delle isole pancreatiche. L'incremento conseguente dei livelli di insulina in circolo stimola l'assorbimento di glucosio da parte dei tessuti periferici e inibisce la gluconeogenesi epatica.

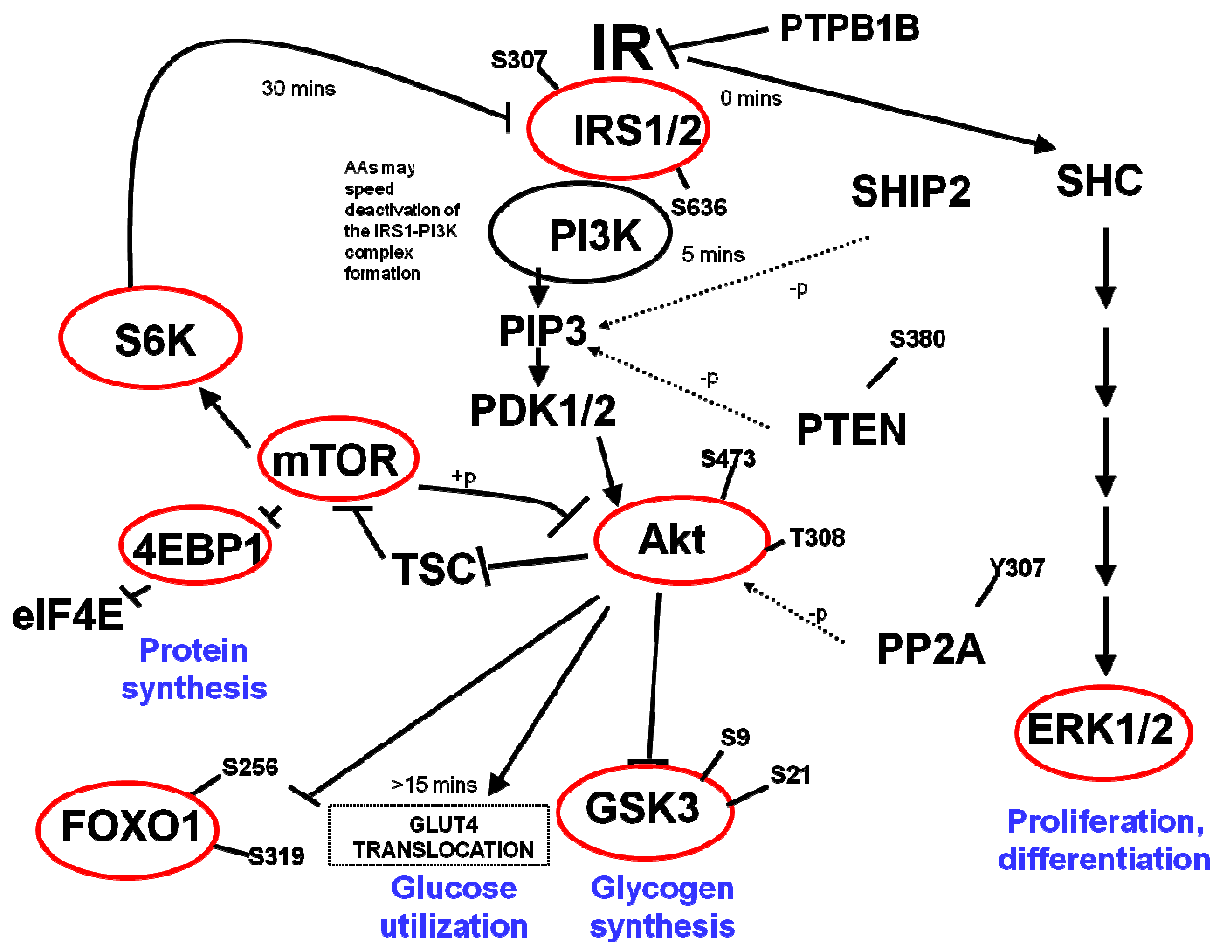


**Figura 0.1. Andamento della concentrazione giornaliera del glucosio e dell'insulina nel sangue.** Il diagramma idealizzato mostra le fluttuazioni del glucosio e dell'insulina plasmatici nell'uomo durante il corso di una giornata contenente 3 pasti. In aggiunta, viene evidenziato l'effetto di un pranzo ricco di zucchero contro quello di un pasto ricco di amido.

Quando si verificano dei difetti nel sistema di controllo in questione, ne consegue un innalzamento cronico dei livelli di glucosio nel sangue, condizione patologica che viene indicata con il termine di diabete mellito. Esistono 2 principali tipi di diabete mellito: diabete mellito di tipo 1 e diabete mellito di tipo 2. Entrambi derivano da complesse interazioni tra geni e ambiente ma presentano patogenesi distinte. Il diabete di tipo 1 è risultato della distruzione immuno-mediata (attacco immunitario mediato dalle cellule T) delle  $\beta$ -cellule delle isole di Langerhans del pancreas, sito di produzione e secrezione dell'insulina, ed è quindi caratterizzato da una assoluta insulino-deficienza. Generalmente la malattia si verifica durante l'infanzia o l'adolescenza, sebbene essa possa verificarsi a tutte le età. Il diabete di tipo 2, invece, si verifica perché la secrezione di insulina è inadeguata e non riesce a far fronte al prevalente difetto dell'azione insulinica, causando un'iperglicemia. Eccessiva assunzione di calorie, inattività e obesità interpretano tutte un ruolo importante nella patogenesi del diabete di tipo 2, malattia che si verifica con una frequenza che aumenta all'aumentare dell'età ed è rara al di sotto dei 40 anni (sebbene ci siano importanti eccezioni).

In aggiunta ai suoi primari effetti sull'omeostasi del glucosio, l'insulina promuove anche diversi altri eventi cellulari, che includono la regolazione del trasporto di ioni e amminoacidi, del metabolismo dei lipidi, della sintesi del glicogeno, della trascrizione genica e del turnover dell'mRNA, della sintesi e della degradazione proteica e della sintesi del DNA. Perciò le azioni dell'insulina interpretano un ruolo chiave nel normale immagazzinamento degli alimenti ingeriti e nella normale crescita e differenziazione cellulare. Per comprendere completamente gli eventi che inducono all'insulino-resistenza e la patofisiologia dell'insulino-deficienza, è necessario identificare, a livello molecolare, i componenti chiave dell'insulin signaling pathway. L'azione insulinica ha inizio con il legame che essa forma con il corrispondente recettore (insulin receptor (IR)) presente sulla membrana cellulare delle cellule target, e si propaga poi attraverso diversi rami. Lo schema corrispondente è mostrato in figura 0.2. Lo scopo di questo lavoro di tesi è la modellizzazione dei sotto-sistemi elementari che compongono il pathway in questione. Si effettua quindi una partition analysis della rete di signaling, scomponendo la rete globale nei singoli moduli che la compongono. La reazione biochimica fondamentale che interessa le specie considerate, che sono tipicamente proteine, è la fosforilazione (o, equivalentemente, la defosforilazione), che consiste nell'aggiunta di un gruppo fosfato alla molecola di interesse. Considerata quindi una certa specie X, un modulo di signaling è definito come il sistema formato dalle molecole non fosforilate e da quelle fosforilate di X e dalle molecole della specie U che controlla, ossia influenza, la reazione di fosforilazione (o di





nolecole per le quali si

defosforilazione) che interessa X. A sua volta, la specie X può essere coinvolta nel controllo della fosforilazione (o di defosforilazione) di un'altra specie. Per la maggior parte delle specie, la forma attiva (ossia la forma in grado di svolgere l'attività di chinasi) è quella fosforilata (anche se esistono delle eccezioni). Un esempio di modulo di signaling è quello che descrive la relazione AKT-GSK3 $\beta$ , per la quale la proteina AKT, nella forma fosforilata, favorisce la fosforilazione della proteina GSK3 $\beta$  (in questo caso particolare GSK3 $\beta$  risulta attiva nella forma non fosforilata, quindi AKT fosforilata interviene favorendo l'inattivazione di GSK3 $\beta$ ). I modelli utilizzati per rappresentare i moduli sono dei modelli compartimentali, quindi dei modelli matematici basati sul principio del bilancio di massa, e sono essenzialmente 2. Il primo è basato sull'ipotesi che la specie U agisca controllando la fosforilazione di X, mentre il secondo è basato sull'ipotesi che la specie U agisca controllando la defosforilazione di X. Comunque, manipolando le equazioni differenziali che derivano dai 2 modelli citati, introducendo alcune ipotesi ragionevoli e ipotizzando delle equazioni di misura che

esprimano una relazione lineare tra variabili di misura e variabili di stato, si ottiene in entrambi i casi la stessa equazione ingresso/uscita, che dipende da 3 parametri univocamente identificabili a priori. Ciò che cambia nei 2 diversi casi è solamente il significato assunto dai parametri. I dati acquisiti fanno riferimento alle serie temporali della quantità totale e della quantità della forma fosforilata delle diverse molecole, conseguenti all'applicazione di differenti stimoli: insulina, leucina e leucina+insulina. La stima parametrica viene quindi effettuata implementando l'equazione ingresso/uscita in Matlab, fornendo in ingresso la serie temporale (interpolata linearmente) della quantità fosforilata di U e sfruttando il metodo di stima ai minimi quadrati non lineari pesati, attendendo in uscita la serie temporale della quantità fosforilata di X. In base ai dati a disposizione, i sotto-sistemi per i quali viene effettuata l'identificazione dei parametri sono i seguenti: AKT-GSK3 $\beta$ , AKT-mTOR, mTOR-4EBP1, AKT-FOXO1, mTOR-P70S6K. L'obiettivo futuro è senza dubbio quello di considerare un modello globale, che coinvolga tutti i sotto-sistemi presenti nella rete di signaling, in modo da realizzare un'identificazione dei parametri globale, utilizzando come stime iniziali ad esempio quelle ottenute in questo lavoro. In questo modo si potrà ottenere un modello globale opportunamente identificato che potrà essere utilizzato per generare delle curve dose-risposta in silico, che potranno poi essere confrontate con i dati sperimentali in-vitro e in vivo. La struttura del lavoro è la seguente.

Nel capitolo 1 viene effettuata una descrizione della biologia delle molecole coinvolte, che sono generalmente proteine. Vengono quindi inizialmente descritte struttura, produzione e secrezione della molecola di insulina, per poi passare a una discussione dettagliata delle reazioni biochimiche che avvengono all'interno delle cellule target e che sono controllate dall'insulina stessa attraverso l'interazione con il corrispondente recettore.

Nel capitolo 2 vengono illustrati il protocollo e i dati acquisiti, per comprendere quali siano le caratteristiche della configurazione sperimentale e della tecnica di misura scelta, in modo da poter collegare quest'ultime alle caratteristiche delle serie temporali misurate.

Il capitolo 3 è dedicato alla pre-elaborazione dei dati. Inizialmente viene analizzata sia la variabilità tecnica che quella biologica dei dati a disposizione, basandosi su un modello dell'errore a coefficiente di variazione costante. Successivamente vengono presentati 2 differenti metodi di stima dei pattern delle serie temporali, dei quali il primo è basato semplicemente sull'operazione di media, mentre il secondo è costituito da un algoritmo Expectation-Maximization complesso.

Nel capitolo 4 viene presentato il modello adottato. Vengono quindi descritti i modelli compartimentali che possono essere utilizzati per descrivere il singolo modulo di signaling e i

passaggi matematici che conducono all'equazione ingresso/uscita finale, che risulta avere la stessa struttura nei 2 diversi casi che possono essere considerati (controllo sulla fosforilazione o sulla defosforilazione).

Nel capitolo 5 viene riportata la tecnica utilizzata per l'identificazione del modello a partire dai dati, sia dal punto di vista teorico che implementativo.

Il capitolo 6 è dedicato alla presentazione dei risultati di stima parametrica.

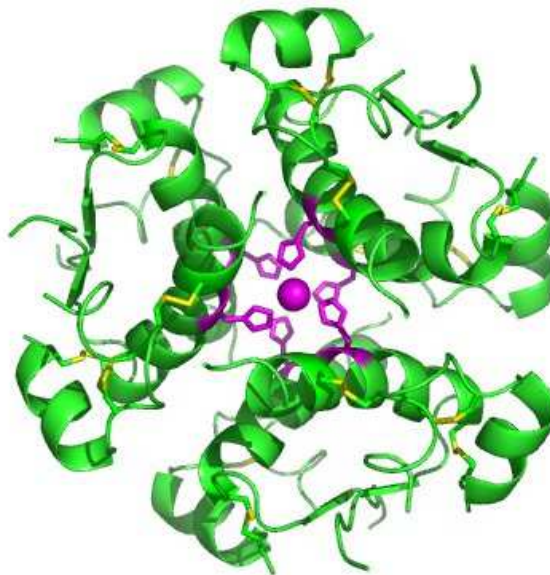
Infine, nel capitolo 7 vengono riportate le conclusioni e gli sviluppi futuri.

# Capitolo 1

## INSULINA E INSULIN SIGNALING PATHWAY

### 1.1 Struttura della molecola di insulina

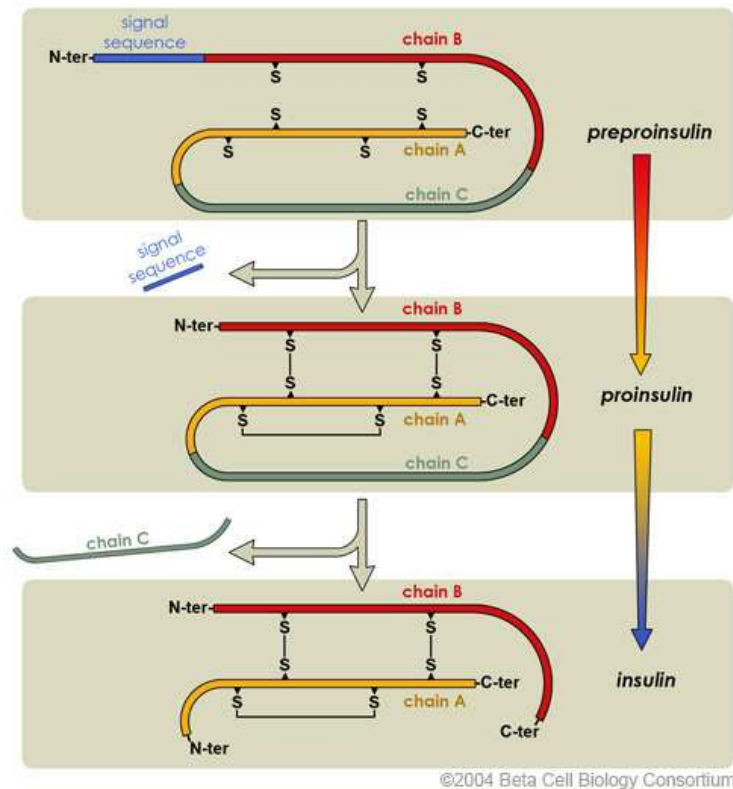
L'insulina umana è una proteina composta da 51 amminoacidi ed ha un peso molecolare di 5808 Da. Essa è costituita da 2 catene polipeptidiche, dette catena A e catena B. Esse sono legate tra loro per mezzo di 2 ponti disolfuro e un ulteriore disolfuro è presente all'interno della catena A. Dei 51 amminoacidi totali, 21 appartengono al peptide A e 30 appartengono al peptide B.



**Figura 1.1. Molecola dell'insulina.** Complesso molecolare formato da 6 molecole di insulina legate tra loro per mezzo di ioni  $Zn^{2+}$  tramite i residui di istidina. Tale forma è quella attraverso la quale la proteina viene immagazzinata nel corpo, mentre la forma attiva è il monomero.

### 1.2 Produzione

Nei mammiferi, l'insulina è sintetizzata nel pancreas all'interno delle  $\beta$ -cellule delle isole di Langerhans (isole pancreatiche). Da 1 a 3 milioni di isole pancreatiche formano la parte endocrina dell'organo, che è principalmente una ghiandola esocrina. La porzione endocrina costituisce solo il 2% della sua massa totale, e all'interno delle isole di Langerhans, le  $\beta$ -cellule rappresentano il 60-80% di tutte le cellule.



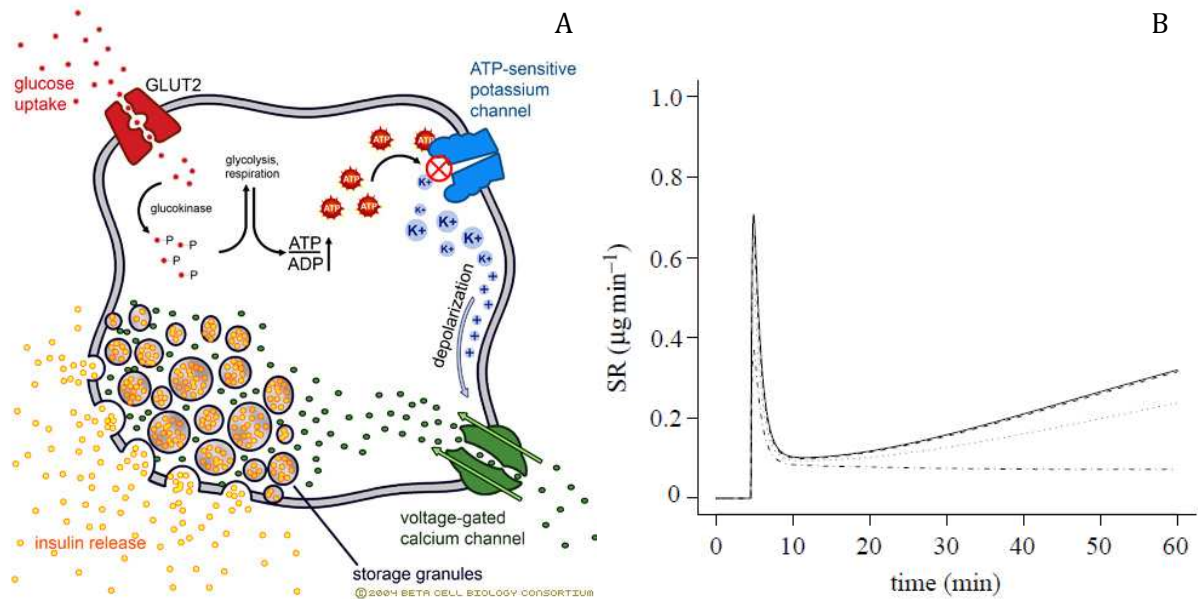
**Figura 1.2. Sintesi dell'insulina.** La produzione di insulina prevede dei passi intermedi. Inizialmente la preproinsulina, che è una molecola inattiva, viene secreta all'interno del reticolo endoplasmatico. L'elaborazione post-traduzionale rimuove la sequenza segnale N-terminale e forma i ponti disolfuro. Alla fine, il polipeptide è tagliato in 2 posizioni per rilasciare la catena C (C-peptide) contenuta. L'insulina attiva risultante viene quindi impacchettata nei granuli secretori per l'immagazzinamento.

L'insulina è una piccola proteina ed è prodotta come parte di una proteina di dimensioni maggiori, con l'obiettivo di permetterle un corretto ripiegamento (folding). Nell'assemblamento proteico dell'insulina, l'mRNA trascritto primario è tradotto in una proteina inattiva detta preproinsulina. La preproinsulina contiene una sequenza segnale ammino-terminale che è richiesta per permettere all'ormone di passare attraverso la membrana del reticolo endoplasmatico (ER) per l'elaborazione post-traduzionale che rimuove quelle componenti non necessarie per l'ormone bioattivo. All'entrata nel reticolo endoplasmatico, la sequenza segnale della preproinsulina, ora inutile, è rimossa proteoliticamente per formare la proinsulina. Una volta verificatasi la formazione post-traduzionale di 3 essenziali legami disolfuro, delle peptidasi specifiche scindono la proinsulina. Il prodotto finale è l'insulina matura e attiva. Dopo essere stata prodotta, essa viene impacchettata e immagazzinata in granuli secretori che vengono accumulati nel citoplasma finché non ne viene fatto scattare il rilascio.

### 1.3 Secrezione

Il processo attraverso il quale l'insulina è rilasciata dalle  $\beta$ -cellule, in risposta a un cambiamento (incremento) della concentrazione del glucosio nel sangue, è un complesso e interessante meccanismo che illustra la natura della regolazione insulinica. I trasportatori del glucosio di tipo 2 (GLUT2) mediano l'entrata del glucosio nelle  $\beta$ -cellule. Come l'alimento grezzo per la glicolisi (pathway universale per la produzione di energia), il glucosio viene fosforilato dall'enzima glucochinasi. Il glucosio modificato viene effettivamente intrappolato all'interno delle  $\beta$ -cellule e viene ulteriormente metabolizzato per produrre ATP, la molecola energetica centrale. L'incremento del rapporto ATP/ADP determina la chiusura dei canali del potassio ATP-dipendenti, impedendo agli ioni potassio di ( $K^+$ ) di fuoriuscire attraverso la membrana cellulare. L'aumento della carica positiva all'interno della cellula, che risulta dall'incremento della concentrazione degli ioni potassio, conduce alla depolarizzazione della cellula. L'effetto netto è l'attivazione dei canali del calcio voltaggio-dipendenti, che trasportano gli ioni calcio ( $Ca^{2+}$ ) all'interno della cellula. Il rapido incremento della concentrazione del calcio fa scattare l'esocitosi dei granuli che contengono l'insulina. Il risultato finale è il rilascio di insulina dalle  $\beta$ -cellule e la sua diffusione all'interno dei vicini vasi sanguigni. L'estensiva vascolarizzazione delle isole pancreatiche assicura una rapida diffusione di insulina (e glucosio) tra  $\beta$ -cellule e vasi sanguigni.

Il rilascio di insulina è un processo bifasico. La quantità iniziale di insulina rilasciata in corrispondenza dell'assorbimento del glucosio è dipendente dalla quantità immagazzinata disponibile. Una volta esaurita tale riserva, ha inizio una seconda fase di rilascio. Questo secondo rilascio è prolungato a partire dal momento in cui l'insulina deve essere sintetizzata, elaborata e secreta per l'intervallo di tempo in cui il livello di glucosio nel sangue rimane sufficientemente elevato da fungere da stimolo per il rilascio. Inoltre, le cellule  $\beta$  devono anche rigenerare la riserva di insulina inizialmente esaurita nella fase di risposta rapida.



**Figura 1.3. Rilascio dell'insulina.** (A) La secrezione di insulina nelle  $\beta$ -cellule è determinata dall'innalzamento dei livelli di glucosio nel sangue. A partire dall'assorbimento di glucosio attraverso i trasportatori GLUT2, la fosforilazione glicolitica del glucosio causa un'incremento del rapporto ATP/ADP che inattiva i canali del potassio, determinando quindi la depolarizzazione della membrana. La depolarizzazione della membrana causa l'apertura dei canali del calcio, permettendo l'ingresso degli ioni calcio dall'esterno. Il conseguente aumento dei livelli del calcio porta al rilascio esocitotico dell'insulina contenuta nei granuli di immagazzinamento. (B) Rilascio insulinico bifasico (simulazione) in risposta a un gradino del livello di glucosio a  $t = 5$  min da  $G = 0$  a  $G = 150 \text{ mg dl}^{-1}$  (curva punto-tratto),  $G = 200 \text{ mg dl}^{-1}$  (curva punto),  $G = 300 \text{ mg dl}^{-1}$  (curva tratto),  $G = 500 \text{ mg dl}^{-1}$  (curva solida).

## 1.4 Insulin signaling pathway

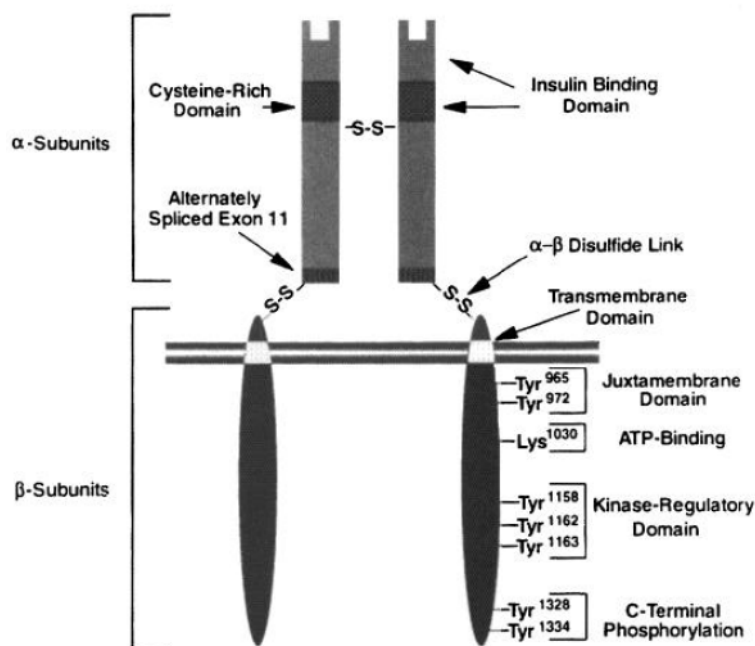
### 1.4.1 Introduzione

L'insulina è in grado di attivare l'assorbimento del glucosio, degli acidi grassi e degli amminoacidi da parte del fegato, del tessuto adiposo e del tessuto muscolare e promuove l'immagazzinamento di questi nutrienti rispettivamente in forma di glicogeno, lipidi e proteine.

### 1.4.2 Componenti del pathway

#### Assorbimento e immagazzinamento del glucosio

Il recettore dell'insulina (IR) è composto di 2 sub-unità  $\alpha$  (dominio extracellulare) e da 2 sub-unità  $\beta$  (dominio transmembrana e intracellulare) legate tra loro per mezzo di ponti disolfuro (Figura 4).



**Figura 1.4. Recettore dell'insulina (IR).** Modello schematico del recettore dell'insulina che mostra i suoi vari domini strutturali e funzionali.

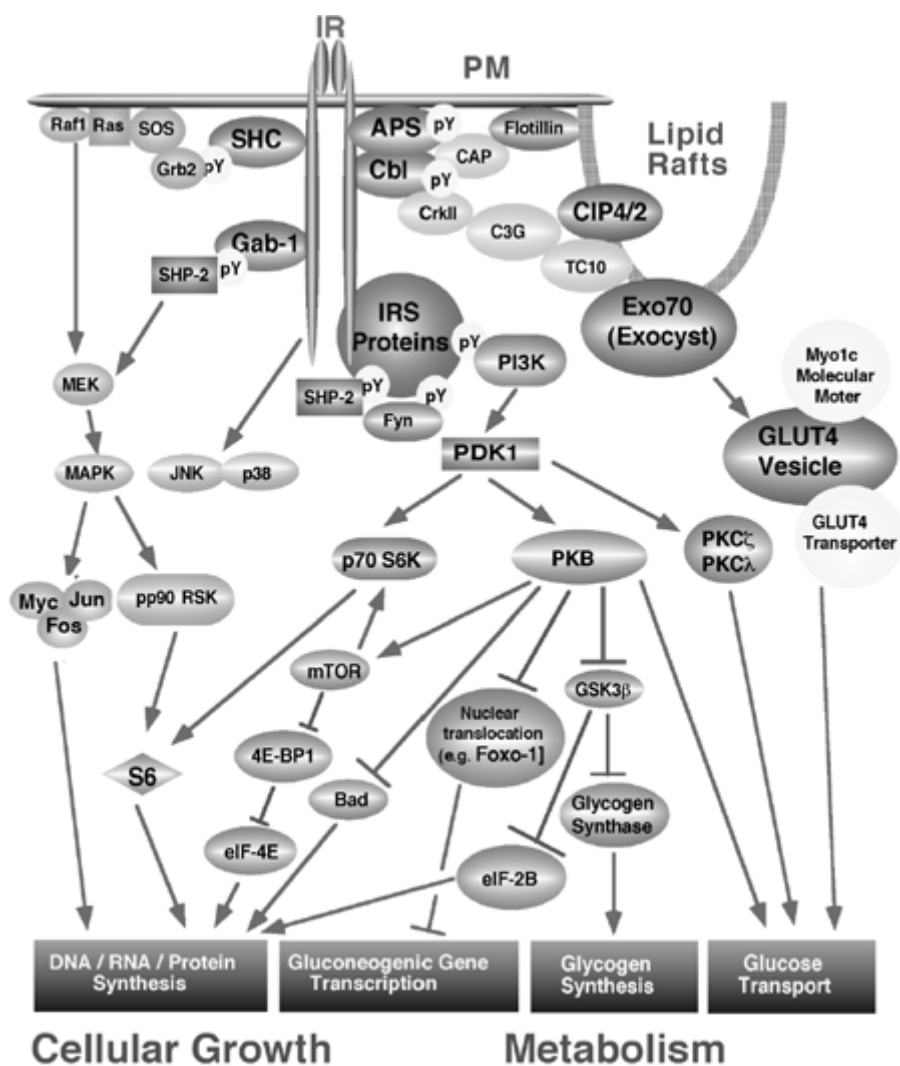
Quando la molecola d'insulina si lega alla sub-unità  $\alpha$  del proprio recettore, quest'ultimo è soggetto a un cambiamento conformazionale che da luogo all'autofosforilazione di diversi residui di tirosina presenti nella sub-unità  $\beta$ . Questi residui sono riconosciuti dai domini fosfotirosina-leganti di proteine adattatore come quelle che appartengono alla famiglia dei substrati del recettore dell'insulina (IRS). L'attivazione del



recettore porta alla fosforilazione di residui di tirosina chiave presenti sulle proteine IRS, come quelli che vengono riconosciuti dal dominio SH2 (Src homology 2) della sub-unità regolatoria p85 della PI3-chinasi (una lipidi-chinasi). In seguito, la subunità catalitica p110 della PI3-chinasi fosforila la molecola PI(4,5)P<sub>2</sub> (phosphatidil-inositol (4,5) biphosphate), portando alla formazione di PI(3,4,5)P<sub>3</sub>. A sua volta, PI(3,4,5)P<sub>3</sub> contribuisce ad attivare la molecola AKT (conosciuta anche come PKB (protein kinase B)), che viene reclutata dalla membrana plasmatica. L'attivazione di AKT richiede anche l'intervento della molecola PDK1 (protein kinase 3-phosphoinositide-dependent protein kinase 1), che in combinazione con una chinasi non ancora identificata porta alla fosforilazione di AKT. Una volta attiva, AKT entra nel citoplasma dove porta alla fosforilazione e quindi all'inattivazione della molecola GSK3 (glucose synthase kinase 3). Il principale substrato di GSK3 è la glicogeno-sintasi (GS), un enzima che catalizza il passo finale nella sintesi del glicogeno. La fosforilazione della glicogeno-sintasi da parte di GSK3 inibisce la sintesi del glicogeno. Perciò l'inattivazione di GSK3 da parte di AKT promuove l'immagazzinamento del glucosio nella forma di glicogeno (omopolimero del glucosio).

Oltre a promuovere l'immagazzinamento del glucosio, l'insulina inibisce la produzione e il rilascio di glucosio dal fegato bloccando la gluconeogenesi e la glicogenolisi. L'insulina controlla direttamente l'attività di un insieme di enzimi metabolici attraverso la fosforilazione e la defosforilazione e regola anche l'espressione di geni che codificano enzimi epatici coinvolti nella gluconeogenesi. Da recenti studi emerge che particolari fattori di trascrizione (definiti forkhead) che sono esclusi dal nucleo in seguito alla fosforilazione da parte di AKT, interpretano un ruolo nella regolazione degli enzimi epatici da parte dell'insulina.

Un'azione chiave dell'insulina riguarda la stimolazione dell'assorbimento del glucosio da parte delle cellule inducendo la traslocazione dei trasportatori del glucosio, GLUT4, dalla riserva intracellulare alla membrana plasmatica, grazie al coinvolgimento delle molecole PI3K e AKT. In aggiunta, un cammino indipendente da PI3K fornisce un secondo stimolo per il reclutamento dei recettori GLUT4 nella membrana plasmatica. In questo cammino, l'attivazione del recettore dell'insulina porta alla fosforilazione della molecola Cbl, che è associata con la proteina adattatore CAP. In seguito alla fosforilazione, il complesso Cbl-CAP trasloca nella membrana plasmatica. Successivamente Cbl interagisce con la proteina adattatore Crk, che è costitutivamente associata al fattore C3G (Rho family guanine nucleotide exchange factor). C3G a sua volta attiva i membri della famiglia di proteine GTP-leganti TC10, che promuovono la traslocazione dei recettori GLUT4 nella membrana plasmatica attraverso l'attivazione di molecole adattatore non ancora conosciute.



**Figura 1.5. Insulin signaling network.** La chinasi attivata del recettore dell'insulina (IRK) fosforila le proteine substrato come Shc, Gab-1, Cbl, APS e le proteine della famiglia dei substrati del recettore dell'insulina (IRS) in corrispondenza di residui di tirosina selettivi che servono come sito di attracco per le molecole successive nel cammino. Questi eventi fanno scattare 3 principali cascate di signaling che hanno origine dalle seguenti chinasi: MAPK (mitogen-activated protein kinase), PI3K (phosphatidylinositol-3 kinase) e Cbl/TC10. Il reclutamento di Grb-2 e mSos a Shc Tyr-fosforilato attiva la cascata di MAPK, mentre l'associazione di PI3K con le proteine IRS determina la produzione di PIP3 (phosphatidyl-inositol (3,4,5) phosphate), che attiva PDK1 (phosphoinositide-dependent kinase-1) e le successive chinasi PKB (protein kinase B), mTOR (mammalian target of rapamycin), p70S6 chinasi, e le tipiche isoforme di PKC $\zeta$ / $\lambda$  (protein kinase C). Il cammino che porta all'attivazione insulino-dipendente di JNK e p38 è poco conosciuto. La fosforilazione di Cbl media il trasporto di glucosio in maniera PI3K-indipendente attraverso l'attivazione di TC10 e la sua interazione con il complesso Exo70 e CIP4/2 (Cdc42-interacting protein 4/2). L'insulina promuove anche il reclutamento di Myoic, che contiene un dominio motore, alle vescicole del trasportatore GLUT-4, facilitando il trasporto del glucosio. Collettivamente, questi cammini mediano le funzioni metabolica e di promozione della crescita possedute dall'insulina, inclusa la traslocazione alla membrana plasmatica delle vescicole contenenti i trasportatori del glucosio GLUT-4, la stimolazione delle sintesi di glicogeno e proteine e l'avvio di una trascrizione genica specifica, inibendo nel coltempo la trascrizione dei geni gluconeogenici.

## **Sintesi proteica**

L'insulina stimola l'assorbimento degli amminoacidi da parte delle cellule, inibisce la degradazione proteica (attraverso un meccanismo non conosciuto) e promuove la sintesi proteica. In condizioni basali, l'attività costitutiva di GSK3 conduce alla fosforilazione e inibizione del fattore eIF2B (guanine nucleotide exchange factor), che regola l'avvio della traduzione proteica. Perciò, in corrispondenza della ricezione di un segnale insulinico, inattivazione di GSK3 da parte di AKT porta alla defosforilazione di eIF2B, promuovendo in tal modo la sintesi proteica e l'immagazzinamento degli amminoacidi. AKT attiva anche la molecola mTOR (mammalian target of rapamycin), che promuove la sintesi proteica attraverso l'attivazione della proteina p70s6k (p70 ribosomal s6 kinase) e l'inibizione della proteina 4E-BP1 (eIF-4E binding protein).

## **Regolazione della sintesi dei lipidi**

L'insulina promuove l'assorbimento degli acidi grassi e la sintesi dei lipidi, mentre inibisce la lipolisi. Recenti studi indicano che la sintesi dei lipidi richiede un incremento della proteina SREBP (transcription factor steroid regulatory element-binding protein). Comunque, il cammino che porta ai cambiamenti nell'espressione di SREBP sono sconosciuti. L'insulina inibisce il metabolismo dei lipidi attraverso il decremento della concentrazione cellulare di cAMP, attivando una fosfodiesterasi specifica per il cAMP negli adipociti.

## **Risposta mitogenica**

Altre proteine trasduttrici di segnale interagiscono con IRS includendo GRB2, una proteina adattatore che contiene domini SH3, che a sua volta si associa con il fattore sos (guanine nucleotide exchange factor son of sevenless) e provoca l'attivazione della cascata MAPK portando alla risposta mitogenica. SHC è un altro substrato per il recettore dell'insulina. In conseguenza della fosforilazione, SHC si associa con GRB2 e può perciò attivare il cammino di MAPK indipendentemente da IRS.

### **1.4.3 Terminazione del pathway**

Enzimi importanti per l'attenuazione del signaling di PI(3,4,5)P3 sono fosfatasi e omologhi della tensina sul cromosoma 10 (PTEN, una 3' fosfatasi) e la famiglia degli omologhi di SRC2 contenente le proteine inositol 5'-fosfatasi (SHIP, una 5'-fosfatasi), che includono i 2 prodotti genici SHIP1 e SHIP2. Queste fosfatasi degradano PI(3,4,5)P3 a PI(4,5)P2 e PI(3,4)P2

rispettivamente. La terminazione del signaling dell'insulina è raggiunto anche grazie all'internalizzazione del complesso insulina-IR all'interno di endosomi e alla degradazione dell'insulina da parte di particolari enzimi (IDE).

## Capitolo 2

### PROTOCOLLO E DATI SPERIMENTALI

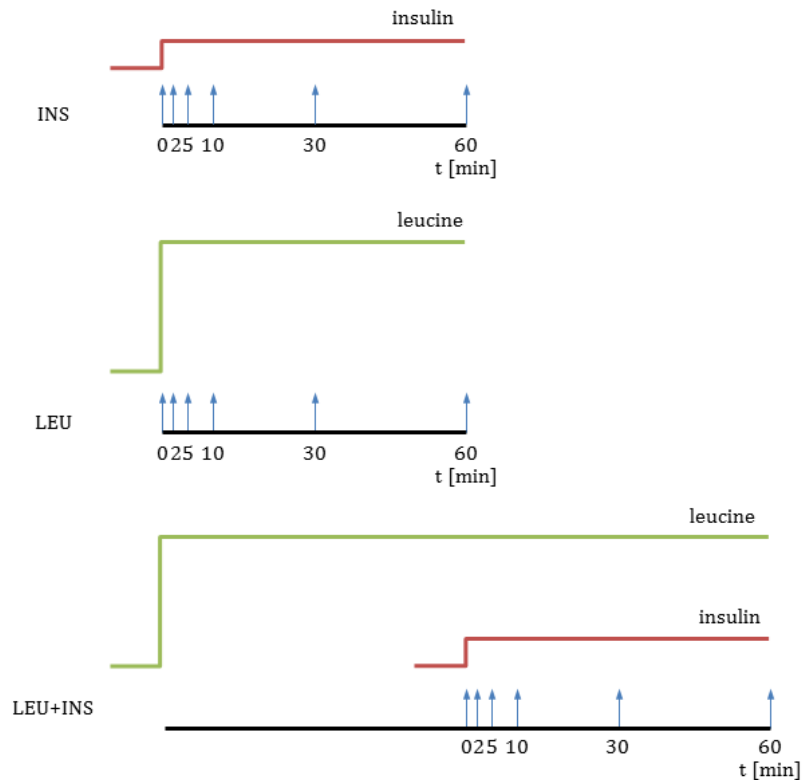
#### 2.1 Configurazione sperimentale

Sono stati effettuati tre esperimenti indipendenti su mioblasti umani di muscolo scheletrico differenziati in miotubuli, utilizzando tre diverse colture cellulari (set A, B e C) come repliche biologiche. Le cellule sono state cresciute a confluenza dell'85% utilizzando DMEM 10% FBS come mezzo di coltura. La notte prima dello stimolo esse sono state trasferite in un mezzo DMEM privo di siero che è stato sostituito con un mezzo privo di amminoacidi (EBSS) un'ora prima dello stimolo. Le cellule sono state raccolte in corrispondenza di 6 punti temporali (0, 2, 5, 10, 30 e 60 minuti) dopo lo stimolo per controllare gli effetti del signaling dell'insulina, sia a breve che a lungo termine. Le linee cellulari sono state esposte a tre differenti stimoli:

1. INS: le cellule sono state stimolate con insulina (EBSS + 100 nM insulina) a  $t = 0$  min
2. LEU: le cellule sono state stimolate con leucina (EBSS + 400  $\mu$ M leucina) a  $t = 0$  min
3. LEU+INS: le cellule sono state pre-incubate con leucina (EBSS + 400  $\mu$ M leucina) per un'ora, per provocare l'insulino-resistenza, e poi stimolate con insulina (EBSS + 400  $\mu$ M leucina + 100 nM insulina)

I livelli di leucina e insulina sono stati controllati per assicurarsi che rimanessero costanti durante l'esperimento.

Un lisato disponibile in commercio è stato caricato in ogni gel come standard interno, ma non si è potuto utilizzare in quanto non ha dato un segnale distinguibile per le proteine target.



**Figura 2.1. Stimoli coinvolti.** I grafici descrivono gli stimoli di INS, LEU e LEU+INS a cui le cellule sono state sottoposte. L'ampiezza dei gradini di concentrazione di insulina è 100 nM, mentre l'ampiezza dei gradini di concentrazione di leucina è 400  $\mu$ M. Le frecce blu indicano gli istanti di campionamento.

## 2.2 Misure

### 2.2.1 Tecnica utilizzata: western blot

La tecnica analitica western blot è utilizzata per rilevare specifiche proteine in un dato campione cellulare. Essa impiega l'elettroforesi su gel per separare le proteine, e la separazione avviene in base alla lunghezza della catena polipeptidica nel caso in cui esse siano in forma denaturata, mentre avviene in base alla struttura tridimensionale nel caso in cui esse siano in forma nativa. Le proteine, dopo essere state separate, vengono trasferite in una membrana (tipicamente costituita da nitrocellulosa o da PVDF) dove possono essere rilevate attraverso l'utilizzo di anticorpi specifici. Di seguito vengono descritti i passi della tecnica.

#### Preparazione del campione

I campioni non possono essere ricavati direttamente dalla cultura cellulare, ma le cellule devono essere rotte attraverso un metodo meccanico. Per favorire la lisi cellulare e la

solubilità delle proteine, possono essere impiegati diversi detergenti, sali e buffer. Spesso vengono aggiunti degli inibitori delle proteasi e delle fosfatasi per evitare la digestione del campione da parte dei propri enzimi. La preparazione del campione avviene generalmente a bassa temperatura per evitare la denaturazione e la degradazione delle proteine.

### **Elettroforesi su gel**

Le proteine del campione vengono separate attraverso l'elettroforesi su gel. La separazione può avvenire in base o al punto isoelettrico (pI), o al peso molecolare, o alla carica elettrica oppure a una combinazione di questi fattori. La natura della separazione dipende dal trattamento del campione e dalla natura del gel. La tipologia più comune di elettroforesi su gel impiega un gel di poliacrilammide e dei buffer caricati con dodecil-solfato di sodio (SDS). Tale tecnica, detta SDS-PAGE, mantiene le catene polipeptidiche in una forma denaturata dal momento che esse sono state trattate con forti agenti riducenti per rimuovere la struttura terziaria e quella secondaria, e perciò ne permette la separazione in base al peso molecolare. Le proteine presentano una carica negativa e migrano verso l'elettrodo positivo attraverso la rete acrilammidica del gel. Le proteine migrano attraverso questa rete con una velocità tanto maggiore quanto minore è la lunghezza della loro catena, permettendone la separazione in base alla dimensione (generalmente misurata in kDa). Nella maggior parte dei blot, la migrazione delle proteine avviene lungo una sola direzione. I campioni vengono caricati nel gel in corrispondenza di appositi pozzetti. Una corsia è generalmente riservata a un marker, ossia a un miscuglio di proteine (disponibile in commercio) aventi pesi molecolari definiti, tipicamente trattate in modo da formare delle bande colorate ben visibili. Quando viene applicata la differenza di potenziale lungo il gel, le proteine migrano al suo interno a differenti velocità e ciò determina la loro separazione in bande all'interno di ogni corsia.

### **Trasferimento**

Al fine di rendere le proteine accessibili ai corrispondenti anticorpi, esse sono trasferite dall'interno del gel alla superficie di una membrana costituita da nitrocellulosa o da PVDF (polivinilidene difluoride). Un metodo di trasferimento che può essere utilizzato, detto elettroblotting, utilizza una corrente elettrica. Esso permette di conservare l'organizzazione delle catene polipeptidiche. In seguito a tale processo, le proteine risultano esposte su una superficie sottile per poter essere rilevate. La membrana utilizzata è scelta per la sua capacità di legare le proteine in maniera non specifica (tutte le proteine sono legate ugualmente bene). Il legame fra membrana e proteine può essere basato sia su interazioni idrofobiche che su interazioni elettrostatiche.

## **Blocking**

Siccome la membrana viene scelta per la sua capacità di legare le proteine e siccome sia gli anticorpi che le molecole target sono proteine, è necessario sottoporre la membrana a un particolare trattamento chimico in modo da bloccare i siti di legame non specifico presenti su di essa, per impedire l'interazione tra gli anticorpi e la membrana stessa. Questo trattamento permette di ridurre il rumore presente nel segnale rilevato, portando a risultati migliori ed eliminando i falsi positivi.

## **Rilevazione**

Durante il processo di rilevazione la membrana viene "sondata" per le proteine di interesse con un anticorpo modificato che risulta legato a un enzima reporter. Quando viene esposto a un appropriato substrato, questo enzima dà luogo a una reazione colorimetrica. Generalmente questo processo viene fatto avvenire in 2 differenti passi.

### **Anticorpo primario**

Gli anticorpi primari sono generati quando una specie ospite o una coltura di cellule immunitarie viene esposta alle proteine di interesse. Generalmente essi costituiscono una parte della risposta immunitaria, mentre nel processo in questione sono raccolti e utilizzati come strumento di rilevazione specifico in grado di legare le proteine direttamente. In tal caso, ad essi viene legato un opportuno marcatore per consentirne la rivelazione. Una soluzione diluita di anticorpo primario viene quindi incubata con la membrana per un periodo di tempo che può variare da 30 minuti a una notte intera.

### **Anticorpo secondario**

In seguito al lavaggio della membrana, necessario per rimuovere l'anticorpo primario non legato, la membrana è esposta ad un altro anticorpo, diretto a una porzione specie-specifica del primario. Gli anticorpi utilizzati provengono da sorgenti animali. L'anticorpo secondario viene generalmente legato a un enzima reporter, che può essere ad es. una perossidasi, e questo permette a diversi anticorpi secondari di legare ogni anticorpo primario, in modo da amplificare il segnale. Generalmente viene utilizzato un anticorpo secondario perossidasi-legante per tagliare l'agente chemiluminescente presente sul primario, e il prodotto di reazione genera un'intensità luminosa proporzionale alla quantità di proteina. Tale segnale luminoso viene quindi rilevato attraverso una CCD camera, in modo da ottenere un'immagine digitale del western blot. L'immagine viene quindi analizzata da un densitometro, che permette di valutare la quantità di proteina marcata.



### 2.2.2 Proteine misurate

Le proteine misurate fino ad ora sono le seguenti: 4EBP1, AKT, ERK12, GSK3 $\beta$ , mTOR, P70S6K, FOXO1 (e  $\beta$ -ACTIN come housekeeping).

Per quanto riguarda IRS1, per ottenere delle bande visibili è necessaria l'immunoprecipitazione: i risultati non sono ancora disponibili. Per le varie proteine considerate, è stata misurata sia la quantità totale che la quantità della forma fosforilata.

<b>proteina totale</b>	<b>proteina fosforilata</b>
4EBP1	p4EBP1 (T37, T46)
AKT	pAKT (S473)
ERK12	ppERK12 (T202, Y204)
GSK3 $\beta$	pGSK3B (S9)
mTOR	pmTOR (S2448)
P70S6K	pP70S6K (T389)
FOXO1	pFOXO1 (S256)

**Figura 2.2. Proteine misurate.** La tabella elenca le proteine misurate e le corrispondenti forme fosforilate (tra parentesi è descritto il sito di fosforilazione).

### 2.2.3 Repliche disponibili

Per ogni replica biologica (set A, B e C) di ogni proteina, le serie temporali di tutti e tre gli stimoli sono state generalmente misurate sullo stesso blot. Le misure sono poi state ripetute in differenti giorni e differenti blot per controllare la riproducibilità tecnica. La tabella seguente (Figura 2.3) riassume la corrispondenza tra blot e repliche tecniche disponibili finora per ogni forma (totale e fosforilata) di ogni proteina, per ogni replica biologica e per ogni stimolo. L'intensità delle bande è stata analizzata utilizzando l'Odyssey infrared image system della LiCor. Questo sistema, basato sulla rilevazione dell'infrarosso, permette di sondare contemporaneamente il lisato con gli anticorpi sia per la quantità di proteina totale che per quella fosforilata, grazie all'utilizzo di due anticorpi secondari marcati con colori differenti. La proteina housekeeping  $\beta$ -ACTIN è stata misurata su una linea separata di alcuni blot e utilizzata come carico di controllo.

(A)

			4EBP1		AKT		ERK12		GSK3B		mTOR		P70S6K		FOXO1		Bacti	
			tot	ph	tot	ph	tot	ph	tot	ph	tot	ph	tot	ph	tot	ph		
INS	A	TR1	D	D	A	A	A	A	A	A	A	A	K	A	K		A	
		TR2	H	H	K	K	H	H	K		D	D		D			D	
		TR3									H	H					H	
	B	TR1	D	D	B	B	B	B	B	B	B	B	B	B	B		L	B
		TR2			L	L	I	I	L		I	I		L				I
		TR3							L		L	L						
	C	TR1	C	C	C	C	J	J	G	G	C	C		M	M	G	C	
		TR2	J	J	M	M			M		G	G				M	J	
		TR3			M	M			M		J	J						
LEU	A	TR1	E	E	A	A	A	A	A	A	A	K	A	K		A		
		TR2	H	H	K		H	H	K		E	E		E			E	
		TR3									H	H					H	
	B	TR1	E	E	B	B	B	B	B	B	B	B	B	B	B			B
		TR2							L		E	E						E
		TR3									L	L						
	C	TR1	C	C	C	C	J	J	G	G	C	C		M	M	G	C	
		TR2	J	J	M	M			M		G	G				M	J	
		TR3			M	M			M		J	J						
LEU+INS	A	TR1	D	D	A	A	A	A	A	A	A	K	A	K		A		
		TR2	H	H	K	K	H	H	K		D	D		D			D	
		TR3									H	H					H	
	B	TR1	D	D	B	B	B	B	B	B	B	B	B	B	B		L	B
		TR2			L	L	I	I	L		I	I		L				I
		TR3							L		L	L						
	C	TR1	C	C	C	C	J	J	G	G	C	C		M	M	G	C	
		TR2	J	J	M	M			M		G	G				M	J	
		TR3			M	M			M		J	J						

(B)

<p>May 18 - May 31</p> <p>BLOT A</p> <p>BLOT B</p>	<p>concentrated lysate Blot 1 (all 3 stimuli for biological replicate A)</p> <p>concentrated lysate Blot 2 (all 3 stimuli for biological replicate B)</p>
<p>Jun 6 - Jun 24</p> <p>BLOT C</p> <p>BLOT D</p> <p>BLOT E</p> <p>BLOT E</p> <p>BLOT G</p>	<p>Blot 1 unconcentrated lysate Set C ins and leu+ins stimuli</p> <p>Blot 2 unconcentrated lysate Set A ins and leu+ins stimuli</p> <p>Blot 3 unconcentrated lysate Sets A and B leu stimuli</p> <p>Blot 4 not utilizable (bad quality)</p> <p>Blot 5 concentrated lysate Set C all 3 stimuli</p>
<p>Jun 27 - Jul 8</p> <p>BLOT H</p> <p>BLOT I</p> <p>BLOT J</p>	<p>Blot 1 lysate Set A (unconcentrated)</p> <p>Blot 2 lysate Set B (unconcentrated)</p> <p>Blot 3 lysate Set C (unconcentrated)</p>
<p>Jul 11 - Aug 25</p> <p>BLOT K</p> <p>BLOT L</p> <p>BLOT M</p>	<p>Blot 1 Set A</p> <p>Blot 2 Set B</p> <p>Blot 3 Set C</p>

Figura 2.3. Corrispondenze tra blot e repliche tecniche. (A) Tabella. (B) Legenda.

TARGET	A			B			C		
	INS	LEU+INS	LEU	INS	LEU+INS	LEU	INS	LEU+INS	LEU
p4EBP1	2	2	1	1	1	1	2	2	2
4EBP1	2	2	2	1	1	1	2	2	2
ratio	1	1	1	1	0	1	1	0	0
pAKTS473	2	2	2	2	2	1	3	3	3
panAKT	2	2	2	2	2	1	3	3	3
ratio	2	2	2	2	2	1	3	3	3
ppERK12	2	2	2	2	2	1	1	1	1
ERK12	2	2	2	2	2	1	1	1	1
ratio	2	2	2	2	2	1	1	1	1
pGSK3 $\beta$	1	1	1	1	1	1	1	1	1
GSK3 $\beta$	2	2	2	3	3	2	3	3	3
ratio	1	1	1	1	1	1	1	1	1
pmTOR	4	4	4	3	3	3	4	4	4
mTOR	4	4	4	3	3	3	4	4	4
ratio	4	3	4	3	2	3	4	4	4
pP70S6K	2	2	2	2	2	1	1	1	1
P70S6K	1	1	1	1	1	1	0	0	0
ratio	0	0	0	2	2	1	0	0	0
pFOXO1	0	0	0	1	1	0	2	2	2
FOXO1	1	1	1	0	0	0	1	1	1
ratio	0	0	0	0	0	0	1	1	1
$\beta$ -actin	3	3	2	2	2	2	2	2	2

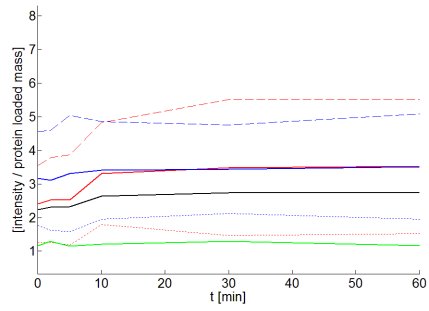
**Figura 2.4. Replicate summary.** I valori indicano quante repliche tecniche sono disponibili per ogni replica biologica (A-C).

#### 2.2.4 Grafici delle serie temporali

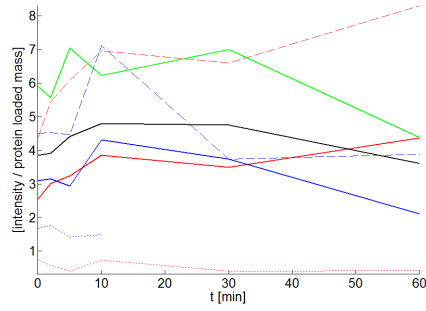
Nei plot seguenti, tre subplot rappresentano i dati per gli stimoli di INS, LEU e LEU+INS rispettivamente. Le repliche tecniche (TR1 in linea tratto, TR2 in linea punto e TR3 in linea tratto-punto) e la loro media (linea continua) sono visualizzate per ogni replica biologica (A, B e C) utilizzando differenti colori (rosso, verde e blu rispettivamente). La linea nera rappresenta la media delle repliche biologiche. I dati visualizzati (e che saranno utilizzati in seguito) sono rapportati alla quantità di lisato caricato nel blot. Tale quantità, nel caso di lisato "standard", è pari a 10  $\mu\text{g}$ , mentre nel caso di lisato concentrato (utilizzato nei casi in cui risulta difficile ricavare le misure) è pari a 89  $\mu\text{g}$ .

### total 4EBP1

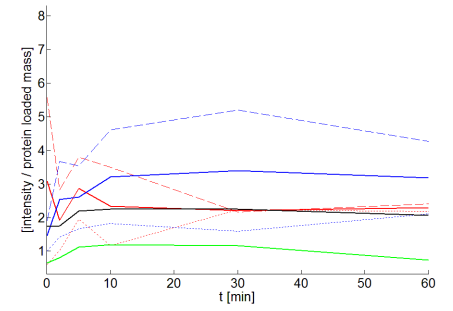
INS



LEU

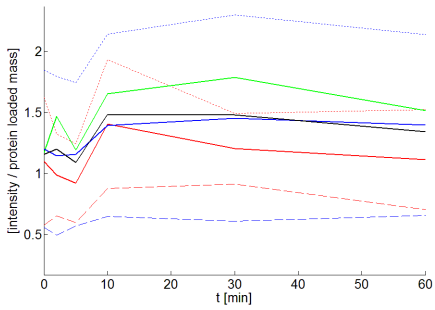


LEU+INS

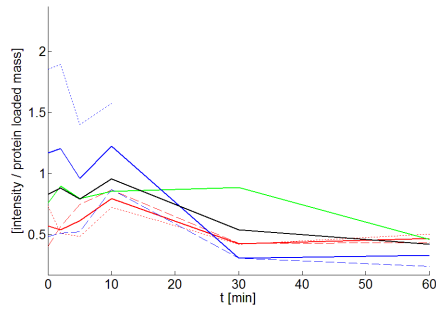


### phosphorylated 4EBP1

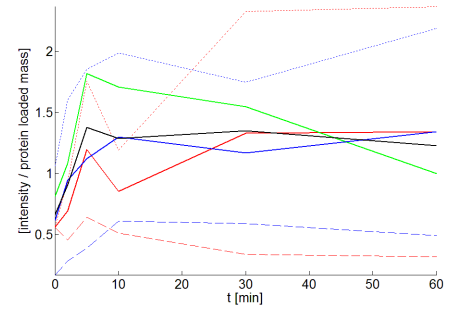
INS



LEU

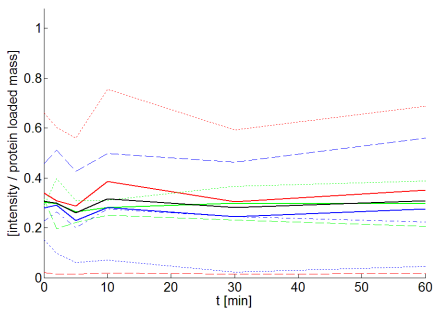


LEU+INS

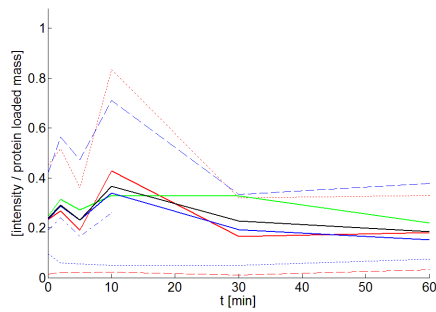


### total AKT

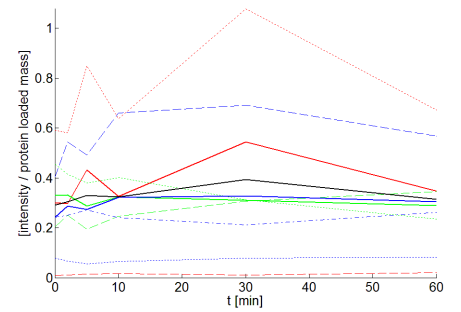
INS



LEU

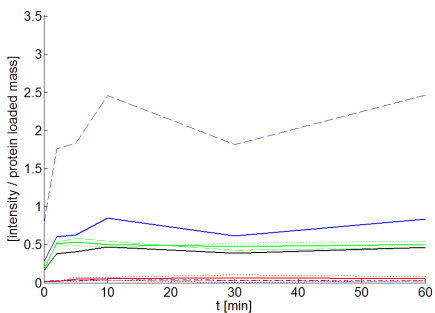


LEU+INS

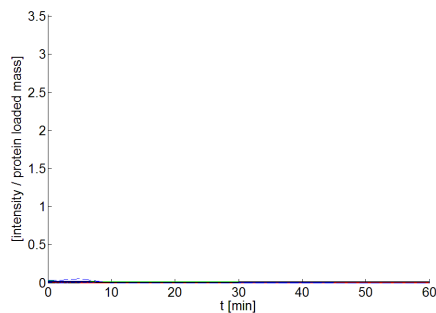


### phosphorylated AKT

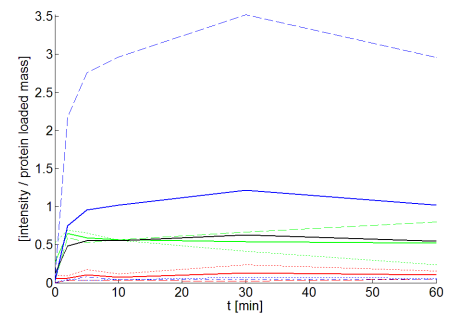
INS



LEU



LEU+INS

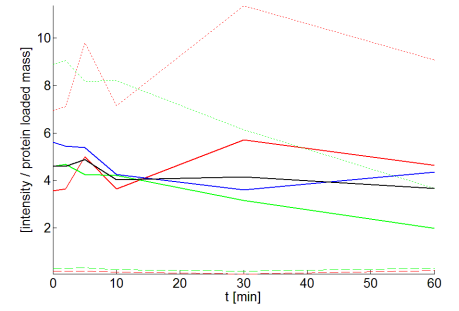
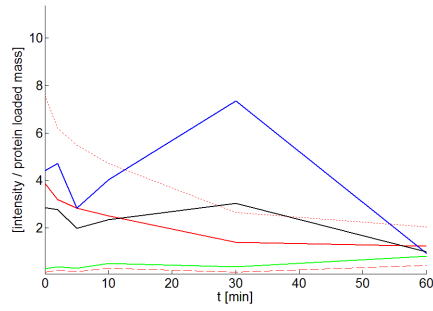
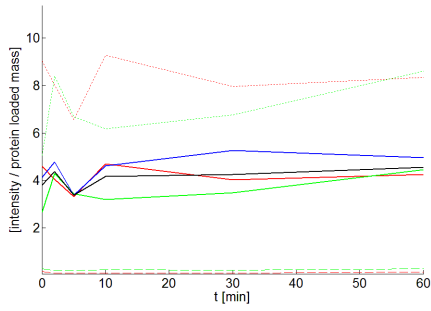


## total ERK12

INS

LEU

LEU+INS

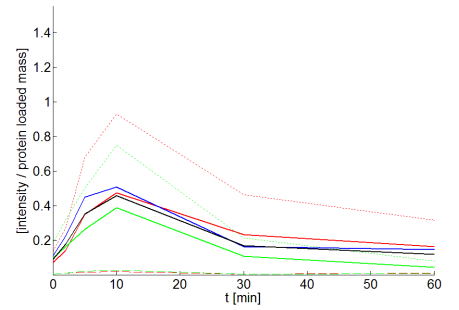
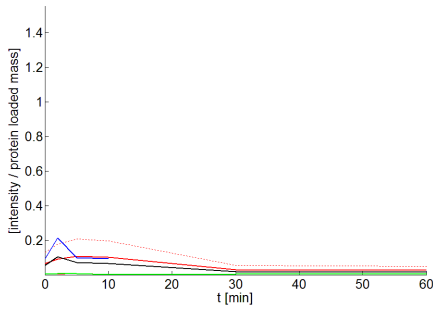
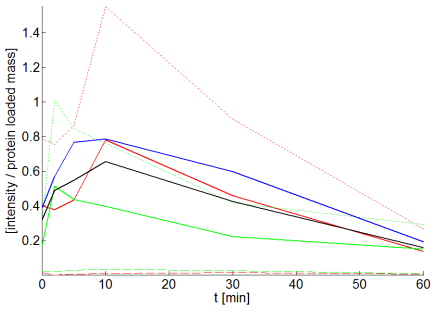


## phosphorylated ERK12

INS

LEU

LEU+INS

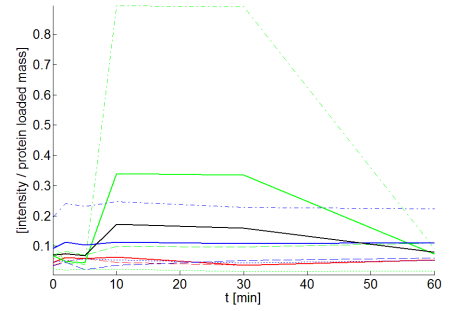
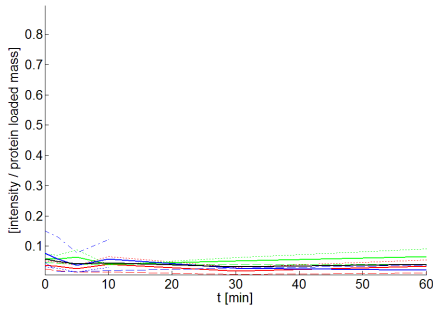
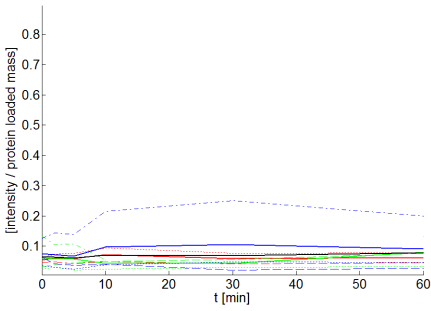


## total GSK3β

INS

LEU

LEU+INS

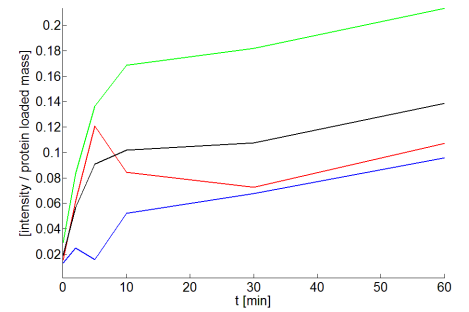
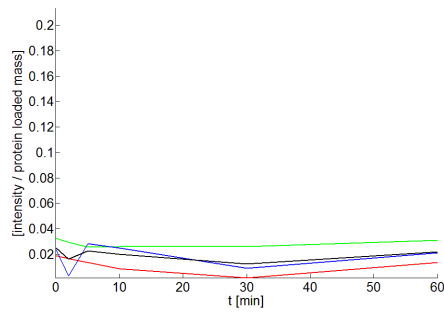
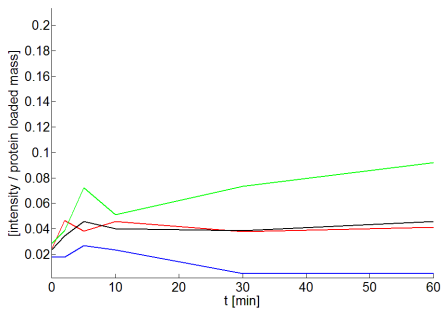


## phosphorylated GSK3β

INS

LEU

LEU+INS

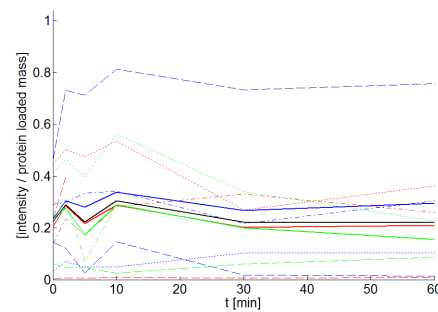
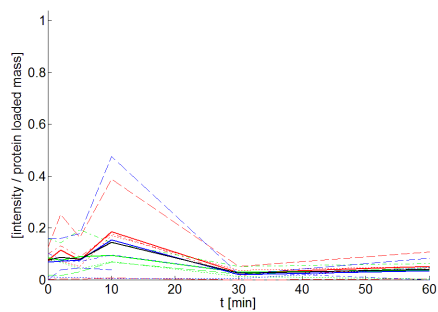
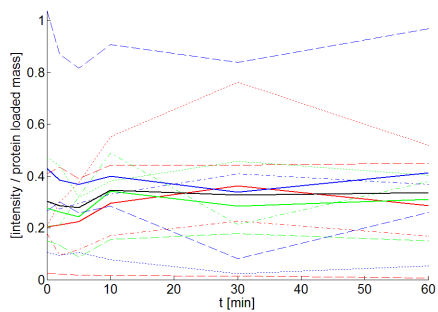


### total mTOR

INS

LEU

LEU+INS

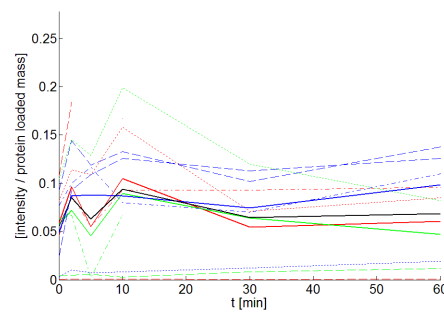
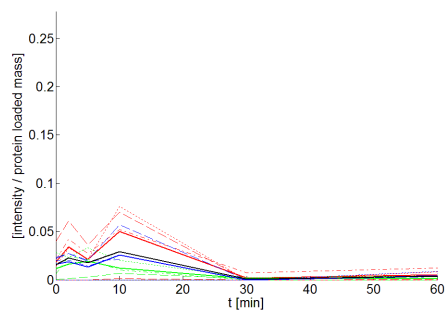
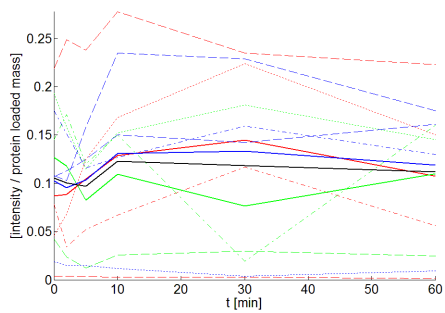


### phosphorylated mTOR

INS

LEU

LEU+INS

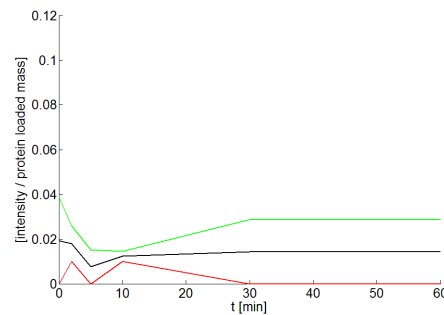
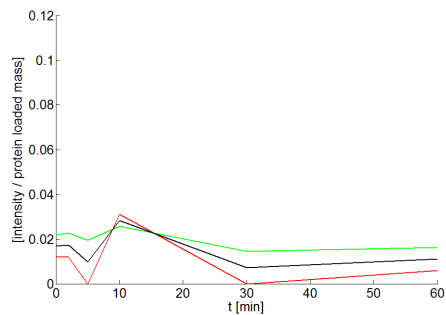
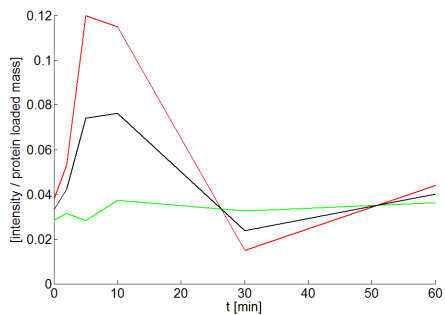


### total P70S6K

INS

LEU

LEU+INS

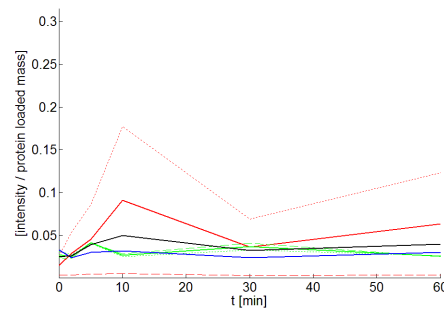
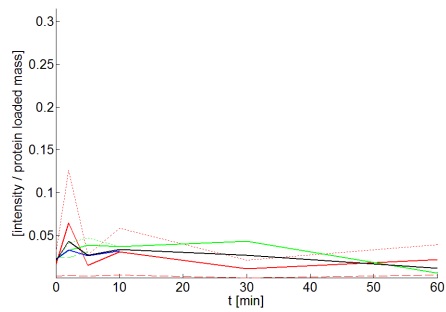
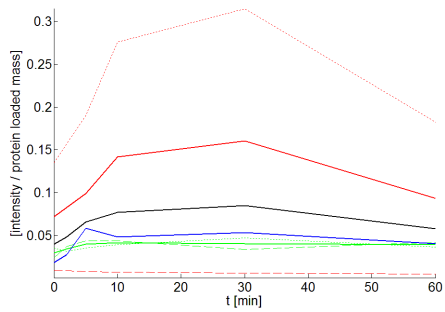


### phosphorylated P70S6K

INS

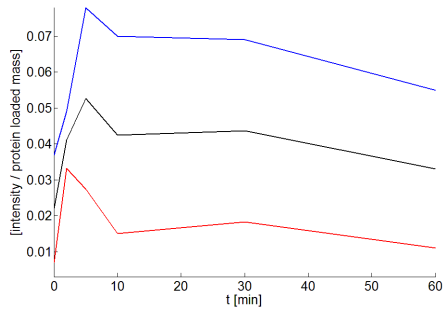
LEU

LEU+INS

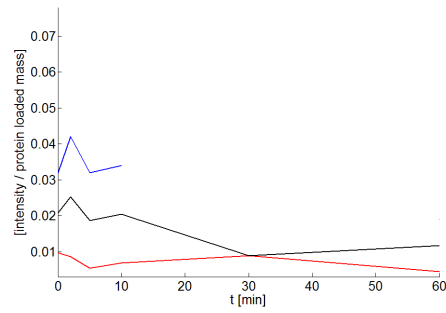


### total FOXO1

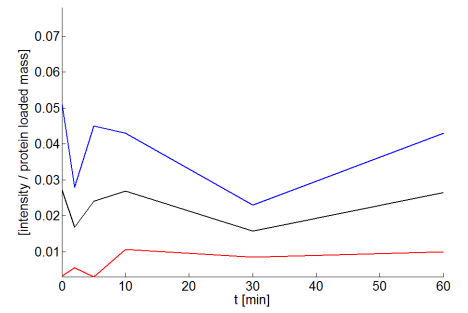
INS



LEU

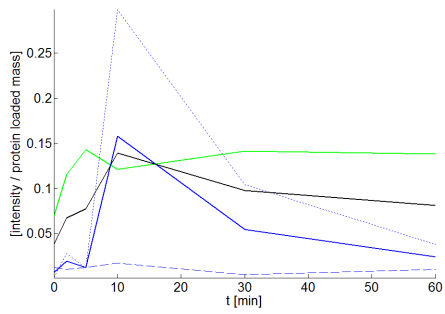


LEU+INS

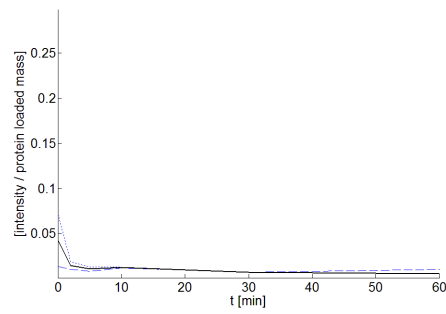


### phosphorylated FOXO1

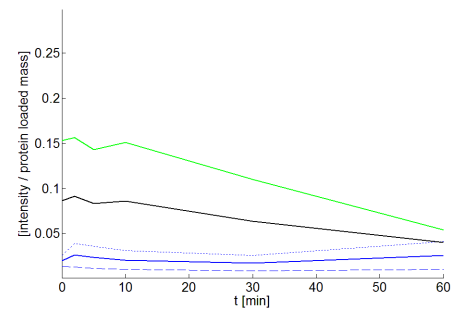
INS



LEU

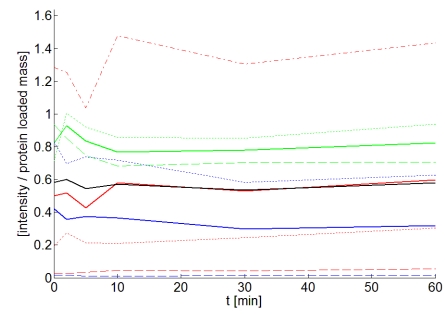


LEU+INS

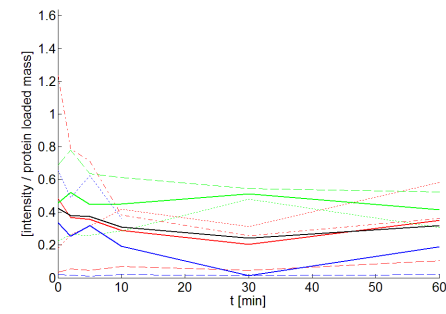


### total $\beta$ -ACTIN (housekeeping)

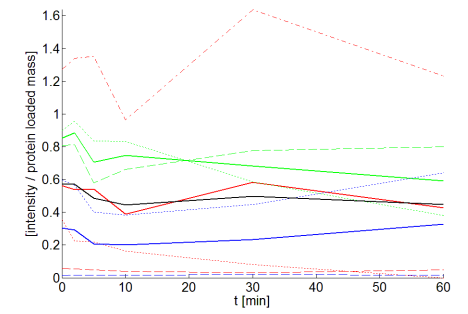
INS



LEU



LEU+INS





## 2.2.5 Valutazione qualitativa dei dati

Osservando i dati acquisiti, si nota che tipicamente le serie temporali corrispondenti alle diverse repliche tecniche di una certa replica biologica presentano uno stesso pattern, ma differiscono tra loro per quanto riguarda l'ampiezza e/o l'offset rispetto al livello zero (preso convenzionalmente come livello di riferimento). Si sa che le misure acquisite sono misure non di concentrazione ma bensì di quantità. In particolare, le misure fanno riferimento a un'intensità luminosa legata alla quantità di proteina presente per mezzo di un fattore dipendente dall'efficienza dell'anticorpo che lega la proteina. La quantità di proteina, a sua volta, dipende dalla concentrazione della proteina stessa e dalla quantità di lisato inserita nel blot per mezzo della pipetta. Per quanto riguarda quest'ultima quantità, nel caso di lisato "standard" essa è pari a 10  $\mu\text{g}$ , mentre nel caso di lisato concentrato (utilizzato nei casi in cui risulta difficile ricavare le misure) essa è pari a 89  $\mu\text{g}$ . Comunque, i valori 10  $\mu\text{g}$  e 89  $\mu\text{g}$  sono puramente teorici, visto che ad essi dovrà essere aggiunto un errore casuale dovuto all'applicazione manuale da parte dell'operatore. Le considerazioni fatte possono quindi giustificare la diversa ampiezza posseduta dalle repliche tecniche considerate (che teoricamente dovrebbero essere identiche), presente sebbene esse siano state preliminarmente rapportate alla quantità di lisato. Per contro, il diverso offset rispetto al livello zero risulta difficilmente giustificabile valutando la configurazione sperimentale. Si ipotizza che esso sia dovuto alla tecnica di misura western blot coinvolta (ad es. esso può essere legato all'intensità luminosa di background).

Considerando che la procedura di stima parametrica viene generalmente condotta su serie temporali "medie", emerge che risulta molto importante valutare quantitativamente la relazione tra le varie repliche coinvolte. Tuttavia tale valutazione non può essere effettuata in maniera soddisfacente considerando direttamente le informazioni di interesse relative alla procedura sperimentale impiegata (come ad es. efficienza degli anticorpi, quantità di lisato, etc.) perché molte di queste informazioni potrebbero essere non ben quantificabili o difficilmente individuabili. A tal proposito, nel capitolo successivo viene descritto un algoritmo Expectation-Maximization di ricerca del pattern che, ricevendo in ingresso tutte le serie temporali disponibili (non distinguendo tra loro le diverse repliche biologiche A, B e C) di una certa proteina (ad es. AKT fosforilata), permette di stimare il pattern comune e i corrispondenti coefficienti moltiplicativi e additivi.

## Capitolo 3

### PRE-ELABORAZIONE DEI DATI

#### 3.1 Modello dell'errore tecnico e biologico

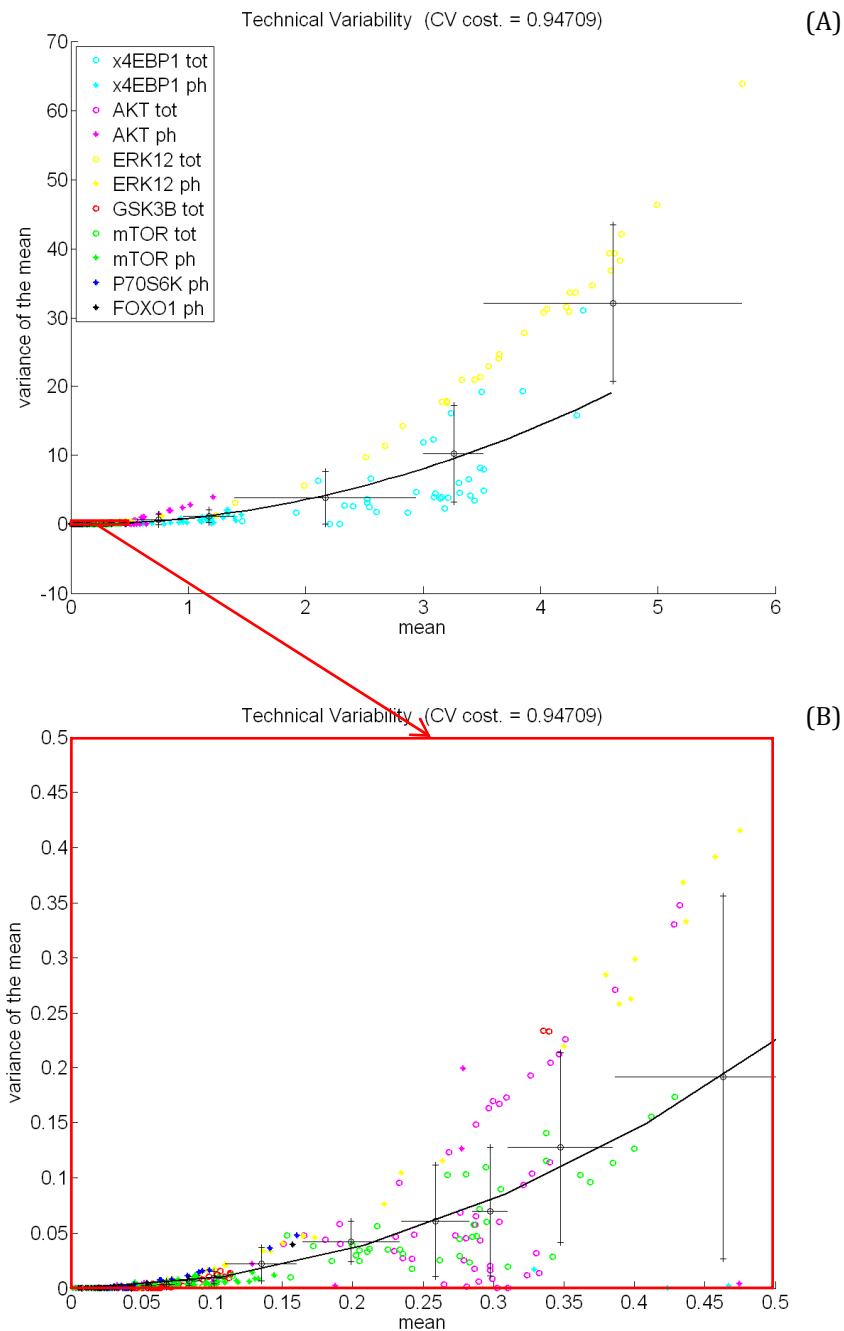
Per quanto riguarda l'errore di misura, si assume che esso abbia una distribuzione gaussiana con media nulla ed una certa varianza che può essere determinata sperimentalmente analizzando le repliche tecniche e biologiche. Ipotizzando che il coefficiente di variazione ( $CV$ ) rimanga costante, il modello per la varianza dell'errore è

$$v_i(t_j) = \beta [z_i(t_j)]^2 \quad (3.1)$$

dove  $\beta$  ( $CV = \sqrt{\beta}$ ) è il parametro che deve essere stimato dalle repliche tracciando la media di ogni replica contro la sua varianza e effettuando il fit del modello su questi dati,  $v_i(t_j)$  indica la varianza dell'errore di misura associato al dato  $z_i(t_j)$  e relativo alla replica  $i$ -esima e all'istante di campionamento  $j$ -esimo. Per definire in maniera appropriata la dipendenza della varianza dall'intensità, il semiasse positivo delle ascisse viene diviso in intervalli e, per ogni intervallo, i valori della varianza vengono mediati.

## Variabilità tecnica

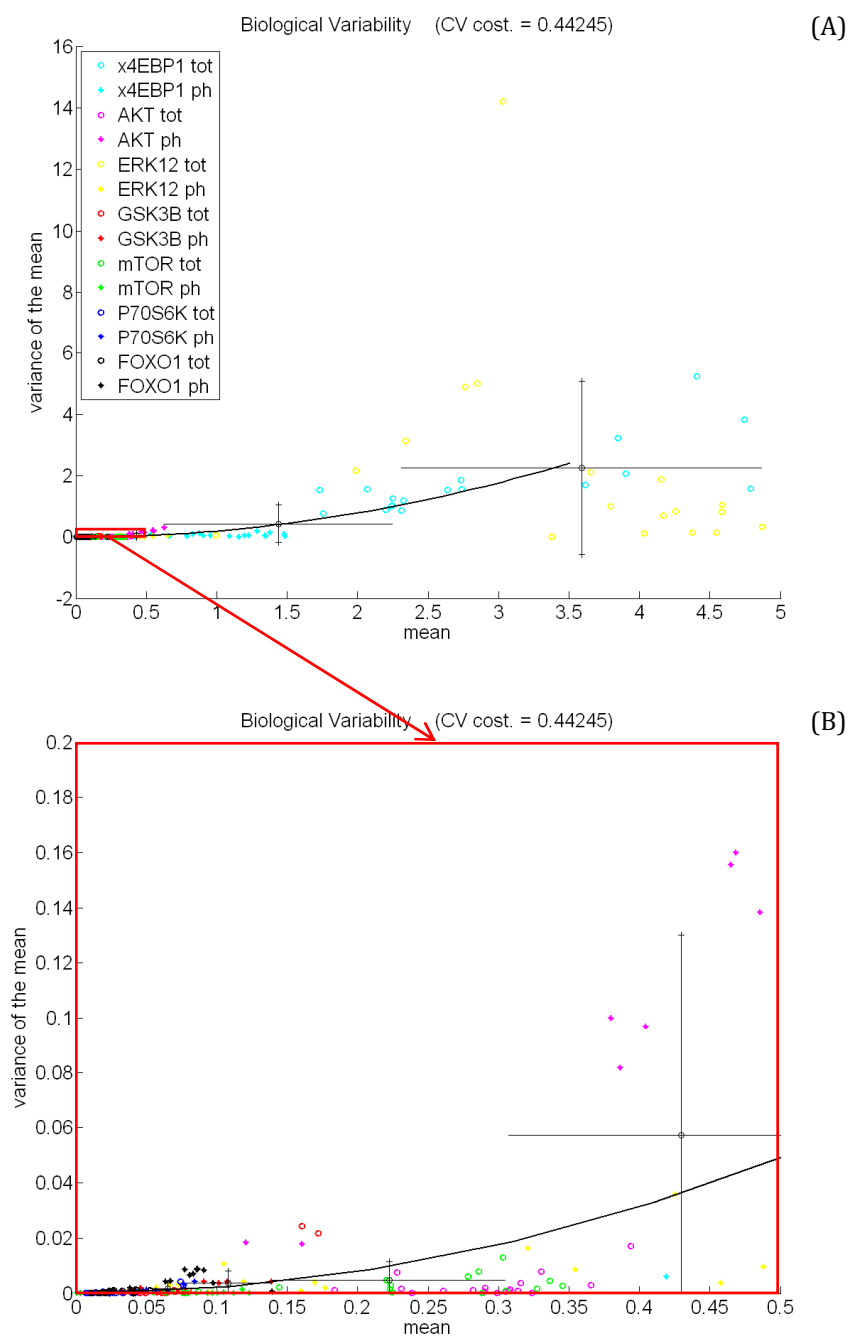
Per quanto riguarda la variabilità tecnica dei dati, il valore stimato per  $\beta$  è 0.90, perciò le repliche tecniche hanno un CV pari all'95%.



**Figura 3.1. Variabilità tecnica delle repliche.** (A) *media vs varianza della media* e relativo fit per tutti i valori e (B) *zoom* relativo all'intervallo [0, 0.5].

## Variabilità biologica

Per quanto riguarda la variabilità biologica dei dati, il valore stimato per  $\beta$  è 0.20, perciò le repliche biologiche hanno un  $CV$  pari al 44%.



**Figura 3.2. Variabilità biologica delle repliche.** (A) *media vs varianza della media* e relativo fit per tutti i valori e (B) *zoom relativo all'intervallo [0, 0.5]*.

### 3.2 Stima del pattern attraverso l'operazione di media

Si ipotizzi che la relazione tra repliche tecniche e corrispondente errore sia descrivibile mediante un modello del tipo

$$y_i(t_j) = P(t_j) + \varepsilon_i(t_j) \quad (3.2)$$

dove con  $i$  si indica la replica e con  $j$  si indica l'istante di campionamento,  $y(t)$  è il segnale misurato,  $P(t)$  è il pattern comune alle diverse repliche considerate,  $\varepsilon(t)$  è l'errore di misura. Assumendo il modello (3.2) e considerando i risultati esposti nel capitolo precedente, emerge che l'errore  $\varepsilon(t)$  risulta piuttosto elevato. In base a ciò, piuttosto che considerare le singole repliche tecniche, è preferibile considerare la loro media (in particolare per quanto riguarda la stima parametrica). Infatti, ipotizzando che tutte le serie temporali disponibili (non distinguendo tra loro le diverse repliche biologiche A, B e C) di una certa proteina condividano lo stesso pattern  $P(t)$  e che l'errore  $\varepsilon(t)$ , inteso come variabile aleatoria, non dipenda dalla replica  $i$  considerata, sia ha che la quantità

$$\bar{y}(t_j) = \sum_{i=1}^{N_R} y_i(t_j) = P(t_j) + \bar{\varepsilon}(t_j) \quad (3.3)$$

che costituisce la media delle  $N_R$  repliche tecniche disponibili, presenta uno scostamento dal pattern  $P(t)$  definito dalle variabili aleatorie  $\bar{\varepsilon}(t_j)$ , le quali sono caratterizzate da deviazioni standard diminuite di un fattore  $\sqrt{N_R}$  rispetto alle variabili  $\varepsilon(t_j)$  d'origine. Attraverso l'operazione di media, quindi, si ottiene una diminuzione dell'entità dell'errore tanto più marcata quanto più è elevato il numero di repliche disponibili.

### 3.3 Stima del pattern attraverso un algoritmo Expectation-Maximization

L'algoritmo Expectation-Maximization che viene qui proposto consiste in una procedura iterativa che, ricevendo in ingresso tutte le serie temporali disponibili (non distinguendo tra loro le diverse repliche biologiche A, B e C) di una certa proteina (ad es. AKT fosforilata), permette di stimare il pattern comune e i corrispondenti coefficienti moltiplicativi e additivi. L'algoritmo è ottimizzato per determinare il pattern corrispondente alle quantità fosforilate, visto che per esse è effettivamente prevista la presenza di un pattern, mentre le quantità totali, in accordo con le (4.7) e (4.20), sono assunte costanti. Comunque la questione non è rilevante, visto che, ai fini della stima dei parametri della (4.30), vengono considerati esclusivamente gli andamenti delle quantità fosforilate.

#### ALGORITMO

---

Fissata la proteina e la forma (ad es. AKT fosforilata), si considera quindi:

$$\text{Modello: } y_i(t_j) = P(t_j)k_i + q_i + \varepsilon_i(t_j) \quad \begin{cases} i = 1, 2, \dots, N_R \\ j = 1, 2, \dots, N_C \end{cases}$$

dove con  $i$  si indica la replica e con  $j$  si indica l'istante di campionamento,  $y(t)$  è il segnale misurato,  $P(t)$  è il pattern comune alle diverse repliche considerate,  $k_i$  e  $q_i$  sono rispettivamente il coefficiente moltiplicativo e il coefficiente additivo associati alla replica  $i$ -esima,  $\varepsilon(t)$  è l'errore di misura,  $N_R$  è il numero di repliche tecniche complessive (considerate le varie repliche biologiche A, B e C) e  $N_C$  è il numero di campioni nella singola replica tecnica ( $N_C = 6$ ).

#### Procedura iterativa:

Fissate la soglia  $CV_{th}$  per i  $CV$  delle stime dei  $k_i$  e dei  $q_i$  e la soglia  $p - value_{th}$  per i  $p - value$  calcolati, si eseguono i seguenti passi:

- 0)  $\hat{P}(t) =$  profilo medio al tempo 0:

$$\hat{P}(t_j) = \frac{1}{N_R} \sum_{i=1}^{N_R} y_i(t_j), \quad j = 1, 2, \dots, N_C$$

- 1)  $\forall i \in \{1, 2, \dots, N_R\}$  si stimano  $k_i$  e  $q_i$  (minimi quadrati):

$$\begin{cases} y_i(t_1) = \hat{P}(t_1)k_i + q_i + \varepsilon_i(t_1) \\ \vdots \\ y_i(t_{N_C}) = \hat{P}(t_{N_C})k_i + q_i + \varepsilon_i(t_{N_C}) \end{cases} \Rightarrow$$

$$\begin{pmatrix} y_i(t_1) \\ \vdots \\ y_i(t_{N_C}) \end{pmatrix} = \begin{pmatrix} \hat{P}(t_1) & 1 \\ \vdots & \vdots \\ \hat{P}(t_{N_C}) & 1 \end{pmatrix} \begin{pmatrix} k_i \\ q_i \end{pmatrix} + \begin{pmatrix} \varepsilon_i(t_1) \\ \vdots \\ \varepsilon_i(t_{N_C}) \end{pmatrix} \Rightarrow$$

$$\mathbf{y}_i = \hat{\mathbf{G}}\mathbf{p}_i + \boldsymbol{\varepsilon}_i$$

$$\hat{\boldsymbol{p}}_i = (\hat{\boldsymbol{G}}^T \boldsymbol{\Sigma}_{\varepsilon_i}^{-1} \hat{\boldsymbol{G}})^{-1} \hat{\boldsymbol{G}}^T \boldsymbol{\Sigma}_{\varepsilon_i}^{-1} \boldsymbol{y}_i$$

dove si ha

$$\left\{ \begin{array}{l} \boldsymbol{\Sigma}_{\varepsilon_i} = \sigma^2 \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \text{ nel caso di } SD = cost \\ \boldsymbol{\Sigma}_{\varepsilon_i} = \sigma^2 \begin{pmatrix} y_i(t_1)^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & y_i(t_{N_C})^2 \end{pmatrix} \text{ nel caso di } CV = cost \end{array} \right. \quad (\sigma^2 \text{ stimata a posteriori}).$$

Successivamente si calcolano i residui pesati per le diverse repliche e i diversi campioni

- per  $SD = cost$ :

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} wres_1(t_1) = y_1(t_1) - [\hat{P}(t_1)\hat{k}_1 + \hat{q}_1] \\ \vdots \\ wres_1(t_{N_C}) = y_1(t_{N_C}) - [\hat{P}(t_{N_C})\hat{k}_1 + \hat{q}_1] \end{array} \right. \\ \left\{ \begin{array}{l} wres_{N_R}(t_1) = y_{N_R}(t_1) - [\hat{P}(t_1)\hat{k}_{N_R} + \hat{q}_{N_R}] \\ \vdots \\ wres_{N_R}(t_{N_C}) = y_{N_R}(t_{N_C}) - [\hat{P}(t_{N_C})\hat{k}_{N_R} + \hat{q}_{N_R}] \end{array} \right. \end{array} \right.$$

- per  $CV = cost$ :

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} wres_1(t_1) = \frac{y_1(t_1) - [\hat{P}(t_1)\hat{k}_1 + \hat{q}_1]}{\sqrt{y_1(t_1)^2}} \\ \vdots \\ wres_1(t_{N_C}) = \frac{y_1(t_{N_C}) - [\hat{P}(t_{N_C})\hat{k}_1 + \hat{q}_1]}{\sqrt{y_1(t_{N_C})^2}} \end{array} \right. \\ \left\{ \begin{array}{l} wres_{N_R}(t_1) = \frac{y_{N_R}(t_1) - [\hat{P}(t_1)\hat{k}_{N_R} + \hat{q}_{N_R}]}{\sqrt{y_{N_R}(t_1)^2}} \\ \vdots \\ wres_{N_R}(t_{N_C}) = \frac{y_{N_R}(t_{N_C}) - [\hat{P}(t_{N_C})\hat{k}_{N_R} + \hat{q}_{N_R}]}{\sqrt{y_{N_R}(t_{N_C})^2}} \end{array} \right. \end{array} \right.$$

e quindi si calcola il  $WRSS$ :

$$WRSS = \sum_{i=1}^{N_R} \sum_{j=1}^{N_C} wres_i(t_j)^2$$

da cui deriva:

$$\hat{\sigma}^2 = \frac{WRSS}{N - M}$$

dove

$$N = N_R N_C \text{ è il numero totale di campioni}$$

$$M = 2N_R + N_C \text{ è il numero totale di parametri } (k_i, q_i \text{ e } P(t_j))$$

e si riscalano i residui per la  $\hat{\sigma}$  trovata:

$$wres_i^{(post)}(t_j) = \frac{wres_i(t_j)}{\hat{\sigma}}$$

Perciò, alla fine, si determinano le deviazioni standard e i  $CV$  delle stime dei  $k_i$  e dei  $q_i$

$$\boldsymbol{Var}[\hat{\boldsymbol{p}}_i] = \boldsymbol{diag}[(\hat{\boldsymbol{G}}^T \boldsymbol{\Sigma}_{\varepsilon_i}^{-1} \hat{\boldsymbol{G}})^{-1}] \text{ (dove in } \boldsymbol{\Sigma}_{\varepsilon_i} \text{ viene considerata la } \hat{\sigma}^2 \text{ stimata)}$$

$$SD[\hat{k}_i] = \sqrt{Var[\hat{k}_i]}, \quad CV[\hat{k}_i] = \frac{SD[\hat{k}_i]}{\hat{k}_i}$$

$$SD[\hat{q}_i] = \sqrt{Var[\hat{q}_i]}, \quad CV[\hat{q}_i] = \frac{SD[\hat{q}_i]}{\hat{q}_i}$$

e il  $p - value_i$

$$v_{Fisher_i} = \frac{WRSS_{media_i} - WRSS_{pred_i}}{WRSS_{pred_i}}$$

$$p - value_i = 1 - F(v_{Fisher_i}, 1, N_C - 2)$$

dove

$$WRSS_{media_i} = (\mathbf{y}_i - \begin{pmatrix} \bar{y}_i \\ \vdots \\ \bar{y}_i \end{pmatrix})^T \boldsymbol{\Sigma}_{\varepsilon_i}^{-1} (\mathbf{y}_i - \begin{pmatrix} \bar{y}_i \\ \vdots \\ \bar{y}_i \end{pmatrix}), \quad \bar{y}_i = \frac{1}{N_C} \sum_{j=1}^{N_C} y_i(t_j)$$

$$WRSS_{pred_i} = (\mathbf{y}_i - \hat{\mathbf{G}}\hat{\mathbf{p}}_i)^T \boldsymbol{\Sigma}_{\varepsilon_i}^{-1} (\mathbf{y}_i - \hat{\mathbf{G}}\hat{\mathbf{p}}_i)$$

$F$  è la funzione di distribuzione di probabilità di Fisher

2) Considerando solo le repliche  $i, \forall i \in \{1, 2, \dots, N_R\}$ , tali da soddisfare le seguenti condizioni:

$$(*) \begin{cases} CV[\hat{k}_i] \leq CV_{th} \\ CV[\hat{q}_i] \leq CV_{th} \\ p - value_i \leq p - value_{th} \end{cases}$$

al tempo  $t_j, \forall j \in \{1, 2, \dots, N_C\}$ , si stima  $P(t_j)$

$$\begin{cases} y_{h_1}(t_j) = P(t_j)\hat{k}_{h_1} + \hat{q}_{h_1} + \varepsilon_{h_1}(t_j) \\ \vdots \\ y_{h_{N_H}}(t_j) = P(t_j)\hat{k}_{h_{N_H}} + \hat{q}_{h_{N_H}} + \varepsilon_{h_{N_H}}(t_j) \end{cases} \Rightarrow$$

$$\begin{pmatrix} y_{h_1}(t_j) \\ \vdots \\ y_{h_{N_H}}(t_j) \end{pmatrix} = \begin{pmatrix} \hat{k}_{h_1} \\ \vdots \\ \hat{k}_{h_{N_H}} \end{pmatrix} P(t_j) + \begin{pmatrix} \hat{q}_{h_1} \\ \vdots \\ \hat{q}_{h_{N_H}} \end{pmatrix} + \begin{pmatrix} \varepsilon_{h_1}(t_j) \\ \vdots \\ \varepsilon_{h_{N_H}}(t_j) \end{pmatrix} \Rightarrow$$

$$\mathbf{y}_j = \hat{\mathbf{k}}P(t_j) + \hat{\mathbf{q}} + \boldsymbol{\varepsilon}_j$$

$$\hat{P}(t_j) = (\hat{\mathbf{k}}^T \boldsymbol{\Sigma}_{\varepsilon_j}^{-1} \hat{\mathbf{k}})^{-1} \hat{\mathbf{k}}^T \boldsymbol{\Sigma}_{\varepsilon_j}^{-1} (\mathbf{y}_j - \hat{\mathbf{q}})$$

dove si ha

$$\left\{ \begin{array}{l} \boldsymbol{\Sigma}_{\varepsilon_j} = \sigma^2 \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} \text{ nel caso di } SD = cost \\ \boldsymbol{\Sigma}_{\varepsilon_j} = \sigma^2 \begin{pmatrix} y_{h_1}(t_j)^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & y_{h_{N_H}}(t_j)^2 \end{pmatrix} \text{ nel caso di } CV = cost \end{array} \right. \quad (\sigma^2 \text{ stimata a posteriori})$$

e i coefficienti  $h_1, \dots, h_{N_H}$  sono i coefficienti delle  $N_H$  repliche che soddisfano le condizioni (\*).

Infine, si determinano le deviazioni standard e i CV delle stime  $\hat{P}(t_j), \forall j \in \{1, 2, \dots, N_C\}$

$$SD[\hat{P}(t_j)] = (\hat{\mathbf{k}}^T \boldsymbol{\Sigma}_{\varepsilon_j}^{-1} \hat{\mathbf{k}})^{-1}$$

$$CV[\hat{P}(t_j)] = \frac{SD[\hat{P}(t_j)]}{\hat{P}(t_j)}$$

3) Si reiterano i punti (1) e (2) finchè non cambiano le stime (a meno di un  $\delta$  piccolo):



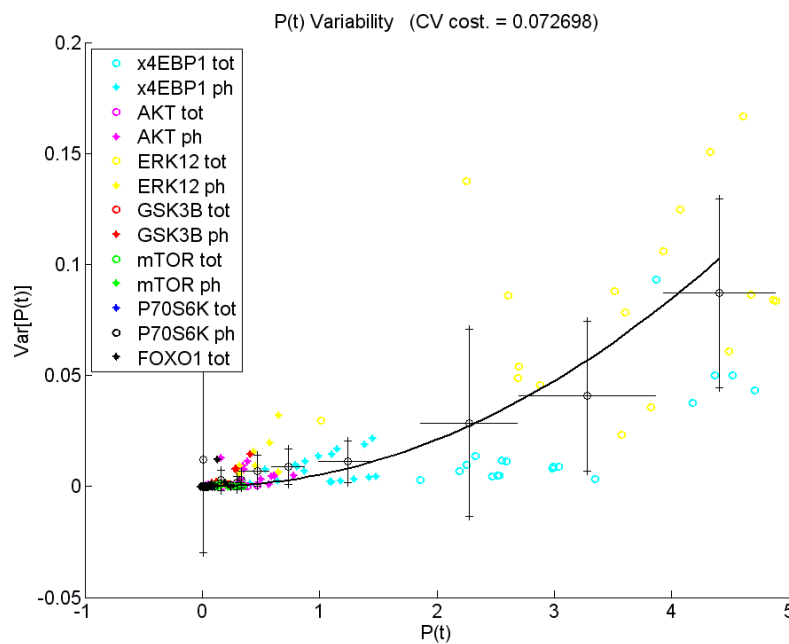
Si reitera fintantochè è soddisfatta la condizione

$$\|\hat{\mathbf{P}}_l - \hat{\mathbf{P}}_{l-1}\| \geq \delta$$

dove si ha

$$\hat{\mathbf{P}}_l = \begin{pmatrix} \hat{P}(t_1) \\ \vdots \\ \hat{P}(t_{N_C}) \end{pmatrix}_l, \text{ vettore dei campioni del pattern } P \text{ all'iterazione } l.$$

L'algoritmo descritto viene implementato in Matlab (versione R2010a), fissando le soglie  $CV_{th} = 300\%$  e  $p - value_{th} = 0.2$ . Esso viene eseguito ipotizzando un modello per l'errore  $\varepsilon(t)$  sia a  $SD = cost$  che a  $CV = cost$ . Il risultato migliore, in particolare per quel che riguarda i  $CV$  (che devono essere bassi), è quello relativo al caso  $CV = cost$ . Una volta ottenuti i risultati per tutte le proteine (anche per le quantità totali, per le quali, dalle condizioni (\*) dell'algoritmo viene eliminata quella relativa al  $p - value$ ), si procede all'analisi dell'errore associato ai pattern  $P(t)$ . Utilizzando ancora il modello d'errore descritto dalla (3.1), si ottiene il risultato in figura 3.3.



**Figura 3.3. Errore associato ai pattern.**

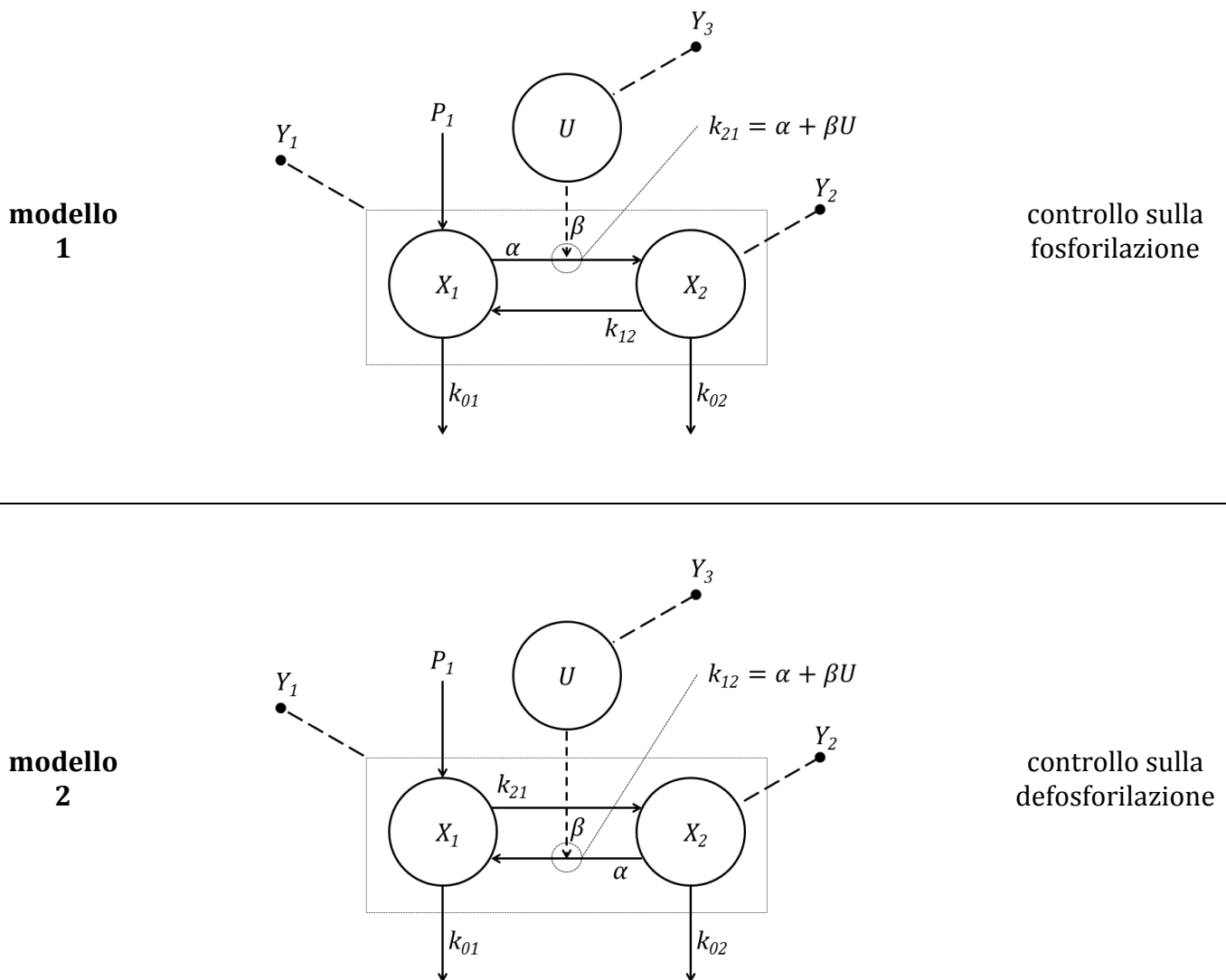
Anche per l'errore associato ai pattern, quindi, il modello a  $CV = cost$  risulta appropriato.

# Capitolo 4

## MODELLI

### 4.1 Modelli di un singolo modulo di signaling

Modelli compartimentali



**Figura 4.1. Modelli di un singolo modulo di signaling.** Modelli compartimentali che descrivono il controllo sulla fosforilazione (modello 1) e il controllo sulla defosforilazione (modello 2).  $U$  è la proteina fosforilata che esercita il controllo,  $X$  è la proteina che viene interessata dalle reazioni chimiche di fosforilazione e defosforilazione ( $X_1$  è la forma non fosforilata e  $X_2$  quella fosforilata),  $P_1$  è la produzione di  $X$ ,  $k_{01}$ ,  $k_{02}$ ,  $k_{12}$  e  $k_{21}$  sono i coefficienti frazionari di trasferimento e  $Y_1$ ,  $Y_2$  e  $Y_3$  sono le misure rispettivamente di  $X_1$ ,  $X_2$  e  $U$ .

## Equazioni di misura

Si ipotizza una relazione lineare tra quantità d'interesse e corrispondente valore misurato:

$$\begin{cases} Y_1 = m_1(X_1 + X_2) + q_1 = m_1 X_{tot} + q_1 & (4.1) \\ Y_2 = m_2 X_2 + q_2 & (4.2) \\ Y_3 = m_3 U + q_3 & (4.3) \end{cases}$$

## Unità di misura

grandezza	unità di misura
$X_1$	quantità
$X_2$	quantità
$U$	quantità
$k_{01}$	$\text{min}^{-1}$
$k_{02}$	$\text{min}^{-1}$
$k_{12}$	$\text{min}^{-1}$
$k_{21}$	$\text{min}^{-1}$
$\alpha$	$\text{min}^{-1}$
$\beta$	$\frac{\text{min}^{-1}}{\text{quantità}}$
$P_1$	quantità · $\text{min}^{-1}$
$Y_1$	$\frac{\text{intensità}}{\text{massa del lisato}}$
$Y_2$	$\frac{\text{intensità}}{\text{massa del lisato}}$
$Y_3$	$\frac{\text{intensità}}{\text{massa del lisato}}$
$m_1$	$\frac{\text{intensità}}{\frac{\text{massa del lisato}}{\text{quantità}}}$
$m_2$	$\frac{\text{intensità}}{\frac{\text{massa del lisato}}{\text{quantità}}}$
$m_3$	$\frac{\text{intensità}}{\frac{\text{massa del lisato}}{\text{quantità}}}$
$q_1$	$\frac{\text{intensità}}{\text{massa del lisato}}$
$q_2$	$\frac{\text{intensità}}{\text{massa del lisato}}$
$q_3$	$\frac{\text{intensità}}{\text{massa del lisato}}$

Figura 4.2. Unità di misura delle grandezze coinvolte.

## 4.2 Equazioni del modello con controllo sulla fosforilazione

- Equazioni di stato

$$\begin{cases} \dot{X}_1 = P_1 - k_{01}X_1 - (\alpha + \beta U)X_1 + k_{12}X_2 & (4.4) \\ \dot{X}_2 = (\alpha + \beta U)X_1 - k_{12}X_2 - k_{02}X_2 & (4.5) \end{cases}$$

- Assunzioni

- $k_{01} = k_{02} = k_0$  (4.6)

- $X_1 + X_2 = X_{tot} = cost$  (4.7)

- $k_{12} \gg k_0$ , quindi  $k_0 + k_{12} \cong k_{12}$  (4.8)

---

- Calcoli

Sommando le equazioni (4.4) e (4.5) e utilizzando la (4.6):

$$(X_1 + X_2) = P_1 - k_0X_1 - k_0X_2 = 0 \quad (4.9)$$

quindi

$$P_1 = k_0(X_1 + X_2). \quad (4.10)$$

Considerando le equazioni (4.4) e (4.5) si ha

$$\begin{cases} \dot{X}_1 = -(\alpha + \beta U)X_1 + (k_0 + k_{12})X_2 & (4.11) \\ \dot{X}_2 = (\alpha + \beta U)X_1 - (k_0 + k_{12})X_2 & (4.12) \end{cases}$$

$$\dot{X}_2 = (\alpha + \beta U)X_1 - (k_0 + k_{12})X_2 \quad (4.12)$$

e siccome dalla (4.7) si ha  $X_1 = X_{tot} - X_2$ ,

l'equazione (4.12) può essere scritta come funzione della sola  $X_2$

$$\dot{X}_2 = (\alpha + \beta U)(X_{tot} - X_2) - (k_0 + k_{12})X_2 \quad (4.13)$$

che, utilizzando la (4.8), porta a

$$\dot{X}_2 = (\alpha + \beta U)(X_{tot} - X_2) - k_{12}X_2. \quad (4.14)$$

In condizioni basali, il sistema si trova in stato stazionario

e quindi l'equazione (4.14) diviene

$$0 = \alpha X_{tot} + \beta U_b X_{tot} - \alpha X_{2b} - \beta U_b X_{2b} - k_{12} X_{2b}. \quad (4.15)$$

Sottraendo l'equazione (4.15) dalla (4.14) si ha

$$\dot{X}_2 = \beta X_{tot}(U - U_b) - (\alpha + k_{12})(X_2 - X_{2b}) - \beta(UX_2 - U_b X_{2b}). \quad (4.16)$$

### 4.3 Equazioni del modello con controllo sulla defosforilazione

- Equazioni di stato

$$\begin{cases} \dot{X}_1 = P_1 - k_{01}X_1 + (\alpha + \beta U)X_2 - k_{21}X_1 & (4.17) \\ \dot{X}_2 = -(\alpha + \beta U)X_2 + k_{21}X_1 - k_{02}X_2 & (4.18) \end{cases}$$

- Assunzioni

- $k_{01} = k_{02} = k_0$  (4.19)

- $X_1 + X_2 = X_{tot} = cost$  (4.20)

- $k_{21} \gg k_0$ , quindi  $k_0 + k_{21} \cong k_{21}$  (4.21)

---

- Calcoli

Sommando le equazioni (4.17) e (4.18) e utilizzando la (4.19):

$$(X_1 + X_2) = P_1 - k_0X_1 - k_0X_2 = 0 \quad (4.22)$$

quindi

$$P_1 = k_0(X_1 + X_2). \quad (4.23)$$

Considerando le equazioni (4.17) e (4.18) si ha

$$\begin{cases} \dot{X}_1 = (\alpha + \beta U)X_2 + k_0X_2 - k_{21}X_1 & (4.24) \\ \dot{X}_2 = -(\alpha + \beta U)X_2 - k_0X_2 + k_{21}X_1 & (4.25) \end{cases}$$

e siccome dalla (4.20) si ha  $X_1 = X_{tot} - X_2$ ,

l'equazione (4.25) può essere scritta come funzione della sola  $X_2$

$$\dot{X}_2 = -(\alpha + k_0 + k_{21})X_2 - \beta UX_2 + k_{21}X_{tot} \quad (4.26)$$

che, utilizzando la (4.21), porta a

$$\dot{X}_2 = -(\alpha + k_{21})X_2 - \beta UX_2 + k_{21}X_{tot} \quad (4.27)$$

In condizioni basali, il sistema si trova in stato stazionario

e quindi l'equazione (4.27) diviene

$$0 = -\alpha X_{2b} - k_{21}X_{2b} - \beta U_b X_{2b} + k_{21}X_{tot}. \quad (4.28)$$

Sottraendo l'equazione (4.28) dalla (4.27) si ha

$$\dot{X}_2 = -(\alpha + k_{21})(X_2 - X_{2b}) - \beta(UX_2 - U_b X_{2b}). \quad (4.29)$$

#### 4.4 Equazione ingresso/uscita del modello

Per entrambi i modelli (modello con controllo sulla fosforilazione e modello con controllo sulla defosforilazione), l'equazione finale che lega l'intensità della proteina fosforilata ( $Y_2$ ) all'intensità della proteina che esercita il controllo ( $Y_3$ ) è la seguente:

$$\dot{Y}_2 = gh(Y_3 - Y_{3b}) - f(Y_2 - Y_{2b}) - g(Y_3 Y_2 - Y_{3b} Y_{2b}), \quad Y_2(0) = Y_{2b} \quad (4.30)$$

dove i parametri  $f$ ,  $g$  e  $h$  sono i parametri univocamente identificabili a priori del modello. Tali parametri sono combinazioni dei parametri originali che compaiono nelle equazioni di stato e nelle equazioni di misura in accordo con la tabella seguente, dove viene considerata anche una situazione più semplice (relazione proporzionale tra variabili di stato e variabili di misura, ossia  $q_1 = q_2 = q_3 = 0$ ):

controllo	equazione di misura	f	g	h
fosforilazione	$Y = mX + q$	$f = (\alpha + k_{12}) - \frac{\beta}{m_3} q_3$	$g = \frac{\beta}{m_3}$	$h = X_{tot} m_2 + q_2$
	$Y = mX$	$f = (\alpha + k_{12})$	$g = \frac{\beta}{m_3}$	$h = X_{tot} m_2$
defosforilazione	$Y = mX + q$	$f = (\alpha + k_{21}) - \frac{\beta}{m_3} q_3$	$g = \frac{\beta}{m_3}$	$h = q_2$
	$Y = mX$	$f = (\alpha + k_{21})$	$g = \frac{\beta}{m_3}$	$h = 0$

Figura 4.3. Significato dei parametri nei vari casi considerati.

## Capitolo 5

### IDENTIFICAZIONE DEI PARAMETRI

#### 5.1 Descrizione teorica: stima ai minimi quadrati pesati

Si consideri un modello di sistema a tempo continuo  $t$  a un ingresso (segnale  $I(t)$ ) e a un'uscita (segnale  $O(t)$ ) e sia  $\mathbf{p}$  il vettore dei corrispondenti parametri univocamente identificabili a priori. L'uscita predetta dal modello può essere espressa nel seguente modo

$$O(t) = f(t, \mathbf{p}) \quad (5.1)$$

dove  $f$  è una funzione che descrive la dipendenza del segnale d'uscita  $O(t)$  dal tempo  $t$ , attraverso la dipendenza dal segnale d'ingresso  $I(t)$ , e dal vettore dei parametri  $\mathbf{p}$ .

Campionando l'uscita del modello in corrispondenza dei tempi di campionamento  $t_k$ ,  $k = 1, 2, \dots, N_c$ , si ottengono  $N_c$  misure, corrotte da errore di misura, descrivibili come

$$Z_k = O_k + \varepsilon_k = f(t_k, \mathbf{p}) + \varepsilon_k \quad k = 1, 2, \dots, N_c \quad (5.2)$$

dove  $\varepsilon_k$ , ossia l'errore che corrompe la  $k$ -esima misura  $Z_k$ , è considerato additivo e assimilabile a una variabile aleatoria. In particolare, per quanto riguarda le variabili aleatorie  $\varepsilon_k$ ,  $k = 1, 2, \dots, N_c$ , si assume che siano tra loro indipendenti e che abbiano media nulla

$$E[\varepsilon_k] = 0 \quad k = 1, 2, \dots, N_c \quad (5.3)$$

e varianza

$$E[\varepsilon_k^2] = \sigma_k^2 \quad k = 1, 2, \dots, N_c \quad (5.4)$$

dove  $\sigma_k^2$  (varianza del processo) può essere costante oppure no. Utilizzando la seguente notazione vettoriale

$$\mathbf{Z} = [Z_1 \dots Z_{N_c}]^T \quad (5.5)$$

$$\mathbf{O} = [O_1 \dots O_{N_c}]^T \quad (5.6)$$

$$\mathbf{F}(\mathbf{p}) = [f(t_1, \mathbf{p}) \dots f(t_{N_c}, \mathbf{p})]^T \quad (5.7)$$

$$\boldsymbol{\varepsilon} = [\varepsilon_1 \dots \varepsilon_{N_c}]^T \quad (5.8)$$

$$\mathbf{E}[\boldsymbol{\varepsilon}] = \mathbf{0} \quad (5.9)$$

$$\mathbf{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \boldsymbol{\Sigma}_\varepsilon \quad (5.10)$$

il problema di stima parametrica può essere formalizzato nel seguente modo:

**Problema di stima parametrica:** determinare una stima  $\hat{\mathbf{p}}$  del vettore  $\mathbf{p}$  dei parametri, noti il vettore  $\mathbf{Z}$  delle misure e la funzione vettoriale  $\mathbf{F}$ .

La matrice di covarianza  $\boldsymbol{\Sigma}_\varepsilon$  può essere espressa come

$$\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \mathbf{B} \quad (5.11)$$

dove la matrice  $\mathbf{B}$  è nota, mentre lo scalare  $\sigma^2$  può essere incognito (stimato a posteriori). Nel caso di varianza costante si ha  $\mathbf{B} = \mathbf{I}_{N_c}$  (identità), mentre nel caso di coefficiente di variazione costante si ha  $\mathbf{B} = \text{diag}([Z_1^2 \dots Z_{N_c}^2]^T)$ .

Nella stima parametrica ai minimi quadrati pesati, la stima  $\hat{\mathbf{p}}$  viene definita come il valore di  $\mathbf{p}$  che, tra tutti i possibili, rende minima la distanza pesata tra modello e dati

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\text{argmin}} [\mathbf{Z} - \mathbf{F}(\mathbf{p})]^T \boldsymbol{\Sigma}_\varepsilon^{-1} [\mathbf{Z} - \mathbf{F}(\mathbf{p})] \quad (5.12)$$

Se il modello è lineare nei parametri, il vettore  $\mathbf{Z}$  delle misure può essere espresso come

$$\mathbf{Z} = \mathbf{G}\mathbf{p} + \boldsymbol{\varepsilon} \quad (5.13)$$

con  $\mathbf{G}$  matrice  $N_c \times N_p$  (dove  $N_p$  è il numero di parametri), e conseguentemente la (5.12) ammette espressione in forma chiusa

$$\hat{\mathbf{p}} = (\mathbf{G}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{G})^{-1} \mathbf{G}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{Z} \quad (5.14)$$

Se il modello è non lineare nei parametri, invece, la (5.12) non è esprimibile in forma chiusa, quindi la stima  $\hat{\mathbf{p}}$  deve essere determinata attraverso l'utilizzo di un metodo iterativo, come ad esempio il metodo del gradiente di Gauss-Newton.

Considerando la (5.11), se  $\sigma^2$  è incognito, in luogo della (5.12) può essere equivalentemente considerata la seguente

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\text{argmin}} [\mathbf{Z} - \mathbf{F}(\mathbf{p})]^T \mathbf{B}^{-1} [\mathbf{Z} - \mathbf{F}(\mathbf{p})] \quad (5.15)$$

e  $\sigma^2$  può essere stimato a posteriori

$$\hat{\sigma}^2 = \frac{WRSS(\hat{\mathbf{p}})}{dof} = \frac{[\mathbf{Z} - \mathbf{F}(\hat{\mathbf{p}})]^T \mathbf{B}^{-1} [\mathbf{Z} - \mathbf{F}(\hat{\mathbf{p}})]}{N_c - N_p} \quad (5.16)$$

Una volta ottenuta la stima  $\hat{\mathbf{p}}$ , occorre valutare la bontà del risultato dell'identificazione attraverso l'analisi dei residui e la valutazione della precisione delle stime.

Per quanto riguarda l'analisi dei residui, ottenuto  $\hat{\mathbf{p}}$  si definisce il vettore *res* dei residui come l'errore di predizione

$$\mathbf{res} = \mathbf{Z} - \mathbf{F}(\hat{\mathbf{p}}) \quad (5.17)$$

Essendo  $\boldsymbol{\varepsilon} = \mathbf{Z} - \mathbf{F}(\mathbf{p})$ , dove  $\mathbf{p}$  è il valore vero, tuttora incognito, si può assumere che i residui costituiscano una stima dell'errore di misura

$$\mathbf{res} = \mathbf{Z} - \mathbf{F}(\hat{\mathbf{p}}) = \hat{\boldsymbol{\varepsilon}}, \quad (5.18)$$

nell'ipotesi che il modello sia corretto (dal punto di vista della struttura e del valore numerico dei parametri), ossia che sia in grado di descrivere correttamente il comportamento del



sistema in questione. Ipotizzando che l'errore di misura sia a campioni scorrelati e siccome si ha

$$\text{Var}\left(\frac{\varepsilon_k}{\sigma_k}\right) = 1, \quad k = 1, 2, \dots, N_c, \quad (5.19)$$

se si definiscono i residui pesati come

$$wres_k = \frac{res_k}{\sigma_k} \quad k = 1, 2, \dots, N_c \quad (5.20)$$

si attende che essi siano poco correlati e per la maggior parte di valore appartenente all'intervallo  $[-1,1]$ . Per la valutazione delle proprietà citate, possono essere utilizzati degli opportuni test statistici nel caso in cui il numero  $N_c$  dei dati a disposizione sia sufficientemente elevato, mentre si ricorre a una semplice ispezione visiva se i dati a disposizione sono pochi.

Per quanto riguarda la precisione delle stime, si ha che la stima di  $\mathbf{p}$  dipende dalle misure  $Z_k$  che sono corrotte da errore (di misura), quindi, a sua volta, la stima  $\hat{\mathbf{p}}$  è affetta da errore (di stima). L'errore di stima è definito come

$$\tilde{\mathbf{p}} = \mathbf{p} - \hat{\mathbf{p}}, \quad (5.21)$$

dove  $\tilde{\mathbf{p}}$  è un vettore aleatorio, essendo aleatorio  $\hat{\mathbf{p}}$ . La variabilità di  $\tilde{\mathbf{p}}$  viene descritta attraverso la corrispondente matrice di covarianza

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{p}}} = \mathbf{Cov}(\tilde{\mathbf{p}}) = \mathbf{E}[\tilde{\mathbf{p}}\tilde{\mathbf{p}}^T], \quad (5.22)$$

che dà un'informazione quantitativa sull'intervallo di valori che l'errore di stima può assumere, fornendo quindi una misura dell'incertezza associata alla stima di  $\mathbf{p}$ . Dalla matrice  $\boldsymbol{\Sigma}_{\tilde{\mathbf{p}}}$  si possono ricavare le deviazioni standard e i coefficienti di variazione delle stime

$$SD(\hat{p}_i) = \sqrt{\text{Var}(\tilde{p}_i)} = \sqrt{\Sigma_{\tilde{p}_{i,i}}}, \quad i = 1, 2, \dots, N_p \quad (5.23)$$

$$CV(\hat{p}_i) = \frac{SD(\hat{p}_i)}{\hat{p}_i} \quad i = 1, 2, \dots, N_p \quad (5.24)$$

Se il modello è lineare nei parametri, la (5.20) può essere espressa in forma esatta

$$\mathbf{Cov}(\tilde{\mathbf{p}}) = (\mathbf{G}^T \boldsymbol{\Sigma}_{\varepsilon}^{-1} \mathbf{G})^{-1}, \quad (5.25)$$

mentre se il modello è non lineare nei parametri, la (5.22) non è esprimibile in forma esatta. In tal caso,  $\mathbf{Cov}(\tilde{\mathbf{p}})$  è approssimabile come

$$\mathbf{Cov}(\tilde{\mathbf{p}}) \cong (\mathbf{S}^T \boldsymbol{\Sigma}_{\varepsilon}^{-1} \mathbf{S})^{-1}, \quad (5.26)$$

dove

$$\mathbf{S} = \begin{pmatrix} \left[ \frac{\partial f(t_1, \mathbf{p})}{\partial p_1} \right]_{\mathbf{p}=\hat{\mathbf{p}}} & \dots & \left[ \frac{\partial f(t_1, \mathbf{p})}{\partial p_{N_P}} \right]_{\mathbf{p}=\hat{\mathbf{p}}} \\ \vdots & \ddots & \vdots \\ \left[ \frac{\partial f(t_{N_C}, \mathbf{p})}{\partial p_1} \right]_{\mathbf{p}=\hat{\mathbf{p}}} & \dots & \left[ \frac{\partial f(t_{N_C}, \mathbf{p})}{\partial p_{N_P}} \right]_{\mathbf{p}=\hat{\mathbf{p}}} \end{pmatrix}. \quad (5.27)$$

## 5.2 Implementazione

La procedura di stima dei parametri  $f$ ,  $g$  e  $h$  della (4.30) viene implementata in Matlab (versione R2010a). La tecnica di stima utilizzata è quella ai minimi quadrati pesati descritta nel paragrafo precedente. La (4.30) non è risolvibile analiticamente, ma è necessario utilizzare un metodo numerico (iterativo) per ottenere l'andamento dell'uscita  $Y_2$  dati l'ingresso  $Y_3$  e il vettore  $\mathbf{p} = [f \ g \ h]^T$  dei parametri, e a tal fine viene utilizzata la function `ode45` con parametro `RelTol = 10^-5` (gli altri parametri vengono mantenuti al valore di default). La custom function sviluppata per implementare la (4.30) è la seguente

```
%Modello 1.
%
%   dY2/dt = g*h*(Y3 - Y3b) - f*(Y2 - Y2b) - g*(Y3*Y2 - Y3b*Y2b)
function dY2=model1(t,Y2,flag,dati)

% parametri
f=dati.p(1);
g=dati.p(2);
h=dati.p(3);

% valori basali
Y3b=dati.Y3b;
Y2b=dati.Y2b;

% interpolazione lineare dell'ingresso
Y3=interp1(dati.t_camp,dati.input,t);

dY2=g*h*(Y3-Y3b)-f*(Y2-Y2b)-g*(Y3*Y2-Y3b*Y2b); % dY2/dt
```

dove i parametri in ingresso sono l'istante di tempo  $t$  e il valore dell'uscita  $Y_2$  attuali nel processo di integrazione e la struttura `dati`, la quale contiene il vettore dei parametri `dati.p`, i valori basali `dati.Y3b` e `dati.Y2b` rispettivamente del segnale d'ingresso e del segnale d'uscita, il vettore dei tempi di campionamento `dati.t_camp` e la serie temporale in ingresso `dati.input`. Siccome il vettore  $\mathbf{Z}$  delle misure non è esprimibile nella forma (5.13), ossia in una forma che esprima una proporzionalità diretta tra  $\mathbf{Z}$  e  $\mathbf{p}$ , la (5.12) non è esprimibile in forma esatta e viene quindi utilizzata la function `lsqnonlin` con parametri di ottimizzazione `TolX = 10^-5`, `TolFun = 10^-5` e `Algorithm = 'trust-region-`

reflective' (gli altri parametri vengono mantenuti al valore di default) per risolvere la (5.12). La custom function che viene passata alla `lsqnonlin` è la seguente

```
%F(p) per il modello 1.
function F=F_model1(p,dati)

% parametri
dati.p=p;

t_camp=dati.t_camp;

opts=odeset('RelTol',1e-5);
[T,Y2]=ode45('modell1',t_camp,dati.output(1),opts,dati);

F=(dati.output-Y2)./dati.sigmas; % F(p)
```

dove i parametri in ingresso sono il vettore dei parametri  $p$  e la struttura `dati`, la quale contiene la serie temporale d'uscita `dati.output` e il vettore delle deviazioni standard `dati.sigmas` associate agli errori  $\varepsilon_k$ ,  $k = 1, 2, \dots, N_c$ , oltre che i parametri che essa contiene quando viene passata alla function `modell1`. Alla luce dei risultati ottenuti nel capitolo 3, sia che si considerino i dati medi che i pattern Expectation-Maximization, il modello dell'errore di misura che si decide di utilizzare è un modello a  $CV$  costante e non noto, stimato a posteriori, e viene quindi risolta la (5.15) in luogo della (5.12). Nella pratica, quello che si fa è porre `dati.sigmas = abs(Z2)`, dove `Z2` è il vettore dei dati d'uscita (ciò equivale a porre un  $CV$  costante pari a 1). La stima viene quindi condotta utilizzando `lsqnonlin` nel seguente modo

```
[p_st,WRSS,wres,exitf,output,Lambda,J]=lsqnonlin('F_model1',p0,lb,ub,opts,dati);
```

dove i parametri in ingresso sono il vettore iniziale dei parametri  $p_0$ , i lower- e gli upper-bounds `lb` e `ub` (`lb = [-Inf 0 0]'` e `ub = [Inf Inf Inf]'` se il modello di misura ipotizzato è  $Y = mX + q$ , mentre `lb = [0 0 0]'` e `ub = [Inf Inf Inf]'` se il modello di misura ipotizzato è  $Y = mX$ ), le opzioni di integrazione `opts` e la struttura `dati`, la quale contiene gli stessi parametri descritti per `F_model1`. Una volta ottenuta le stime, i corrispondenti  $CV$  possono essere ottenuti come

```
covp=inv(J'*J);
CVp=sqrt(diag(covp))./p_st*100;
```

dove, in particolare,  $J$  è lo jacobiano e  $p\_st$  il vettore delle stime restituiti dalla `lsqnonlin`, e la stima di  $\sigma^2$  può essere ottenuta come

$$\text{sigma2\_post} = \text{WRSS} / (\text{NC} - \text{NP});$$

dove  $\text{WRSS}$  è il minimo della funzione costo calcolato in corrispondenza delle stime  $p\_st$  e  $\text{NC}$  e  $\text{NP}$  sono rispettivamente il numero di campioni e il numero di parametri ( $\text{NP} = 3$ ).

Ottenuta la stima  $\text{sigma2\_post}$ , si possono quindi riscalare i  $CV$   $CV_p$  e i residui pesati  $wres$

$$\begin{aligned} \text{covp\_post} &= \text{sigma2\_post} * \text{covp}; \\ \text{CVp\_post} &= \text{sqrt}(\text{diag}(\text{covp\_post})) ./ p\_stim * 100; \\ \\ \text{wres} &= \text{wres} / \text{sqrt}(\text{sigma2\_post}); \end{aligned}$$

Al fine di rendere confrontabili i risultati ottenuti per i diversi casi (al variare del modulo di signaling, dello stimolo e del modello di misura), la ricerca del minimo (assoluto) della funzione costo della (5.15) viene condotta attraverso l'esecuzione di una procedura sistematica che utilizza l'algoritmo genetico (function `ga`) e che realizza un opportuno tuning dei parametri. La procedura utilizzata è descritta attraverso il flow-chart in figura 5.1. L'algoritmo utilizzato ha quindi inizio con l'esecuzione della function `ga` con parametro `PopInitRange = [0; 1]` (gli altri parametri vengono mantenuti al valore di default), al fine di ottenere una prima stima  $\hat{p}_{ga}$  del vettore  $p$ . Quindi  $\hat{p}_{ga}$  viene fornita in ingresso alla function `lsqnonlin`, ottenendo la stima  $\hat{p}_{lsq}$ . Successivamente si effettua un tuning dei parametri, attraverso l'utilizzo della function `lsqnonlin` che permette di ottenere anche la precisione delle stime. Il tuning inizia con delle variazioni relative dei parametri pari a -0.5, 0 e 0.5 e prosegue (eventualmente) iterativamente con delle variazioni relative dei parametri pari a  $-\frac{CV[\hat{p}_i]}{4} \hat{p}_i$ , 0 e  $\frac{CV[\hat{p}_i]}{4} \hat{p}_i$ ,  $i = 1, 2, 3$ , dove  $\hat{p}$  è la stima migliore trovata fino a tal punto, finchè i  $CV$  di  $\hat{p}$  non sono "accettabili". Con il termine "accettabili" s'intende che i  $CV$  abbiano un valore non elevato, e la valutazione viene effettuata soggettivamente senza l'utilizzo di una o più soglie fissate. Un miglioramento futuro per l'algoritmo può essere senza dubbio l'introduzione di un criterio d'arresto oggettivo.

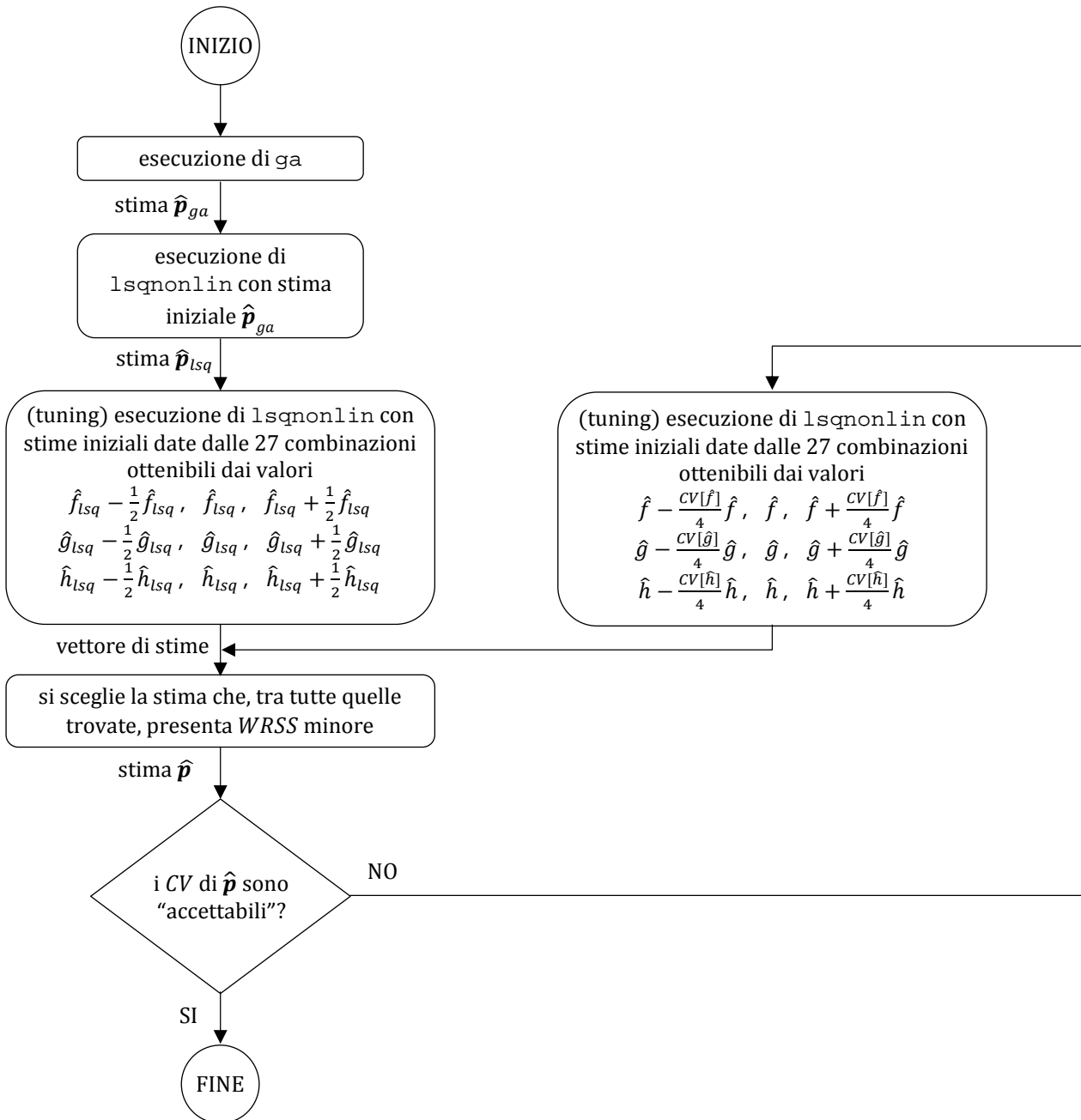


Figura 5.1. Algoritmo di ricerca del minimo della funzione costo.

## Capitolo 6

### RISULTATI

In questo paragrafo vengono presentati i risultati della stima parametrica condotta sia sui dati medi (paragrafo 1), ottenuti in accordo con la (3.3), che sui pattern Expectation-Maximization (paragrafo 2), ottenuti in accordo con l'algoritmo presentato nel paragrafo 1 del capitolo 3. La stima è realizzata attraverso l'implementazione descritta nel paragrafo 2 del capitolo 5. I risultati vengono presentati sia per il modello di misura  $Y = mX + q$  che per il modello di misura  $Y = mX$ .

## 6.1 Dati medi

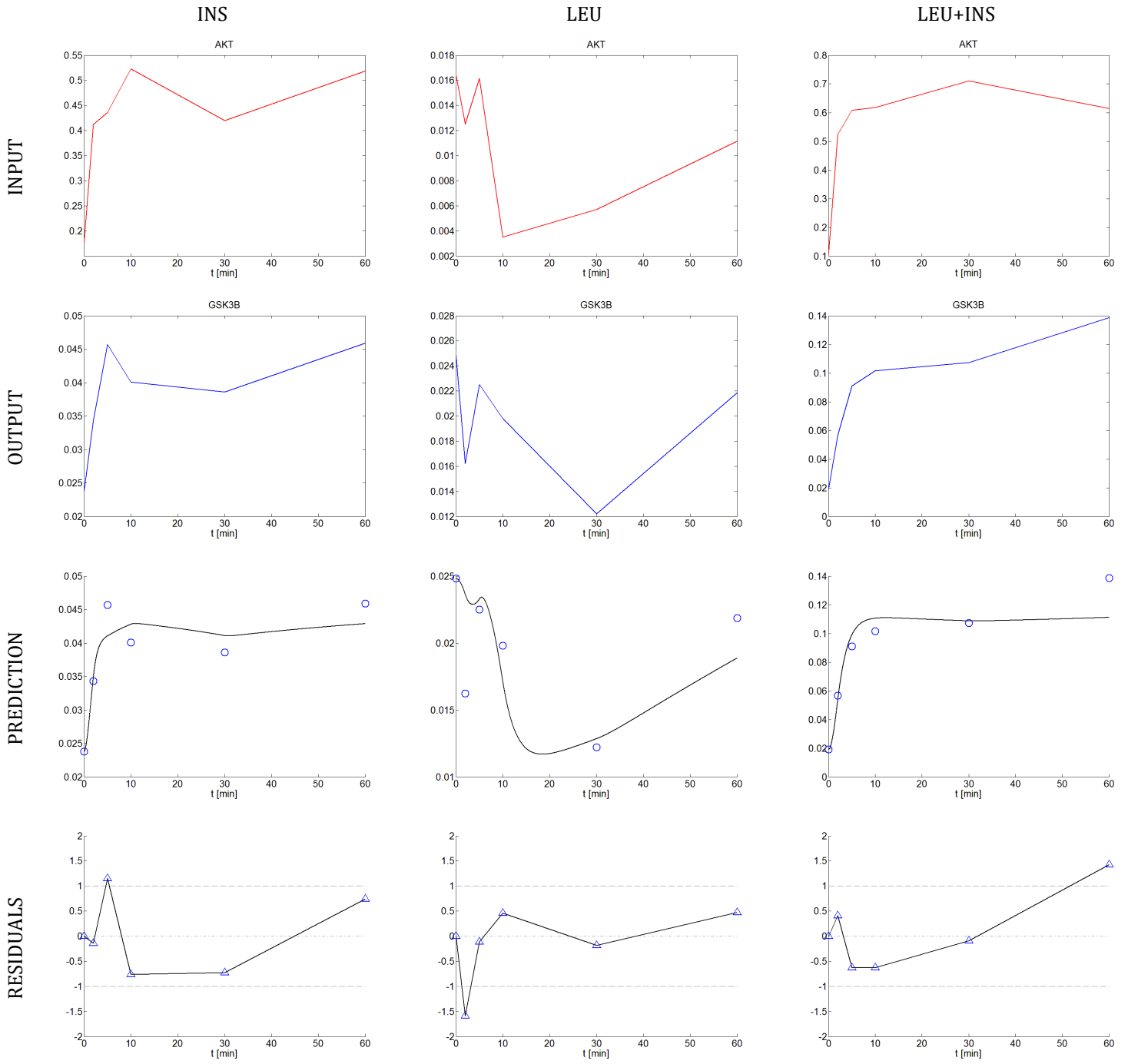
Modello di misura  $Y = mX + q$

Tabella

sotto-sistema	trattamento	F	g	h	WRSS	$\hat{\sigma}$
AKT-GSK3 $\beta$	INS	-0,151 (954%)	3,078 (89%)	0,050 (54%)	0,024	0,089
	LEU	0,324 (178%)	3,504 (1796%)	0,131 (1452%)	0,250	0,289
	LEU+INS	-0,250 (350%)	1,295 (92%)	0,096 (90%)	0,058	0,139
AKT-mTOR	INS	-0,660 (266%)	1,603 (225%)	0,106 (9%)	0,025	0,091
	LEU	0,063 (616%)	0,272 (10989%)	0,400 (10753%)	0,505	0,410
	LEU+INS	-9,660 (8%)	28,647 (18%)	0,065 (7%)	0,072	0,155
mTOR-4EBP1	INS	1,454 (90%)	1,962 (742%)	14,317 (659%)	0,002	0,024
	LEU	0,189 (59%)	11,748 (60%)	1,546 (45%)	0,030	0,100
	LEU+INS	-1,505 (8%)	25,287 (3%)	1,123 (8%)	0,012	0,064
AKT-FOXO1	INS	1,085 (301%)	0,724 (1089%)	0,307 (952%)	0,360	0,347
	LEU	0,054 (1033%)	0,061 (174152%)	4,209 (173749%)	14,018	2,162
	LEU+INS	-0,010 (14495%)	0,019 (11709%)	0,013 (45616%)	0,012	0,062
mTOR-P70S6K	INS	0,538 (917%)	0,429 (10800%)	2,799 (10526%)	0,306	0,319
	LEU	12,619 (2789%)	0,862 (91797%)	6,234 (89463%)	0,461	0,392
	LEU+INS	0,292 (426%)	0,266 (5633%)	1,138 (5537%)	0,065	0,147

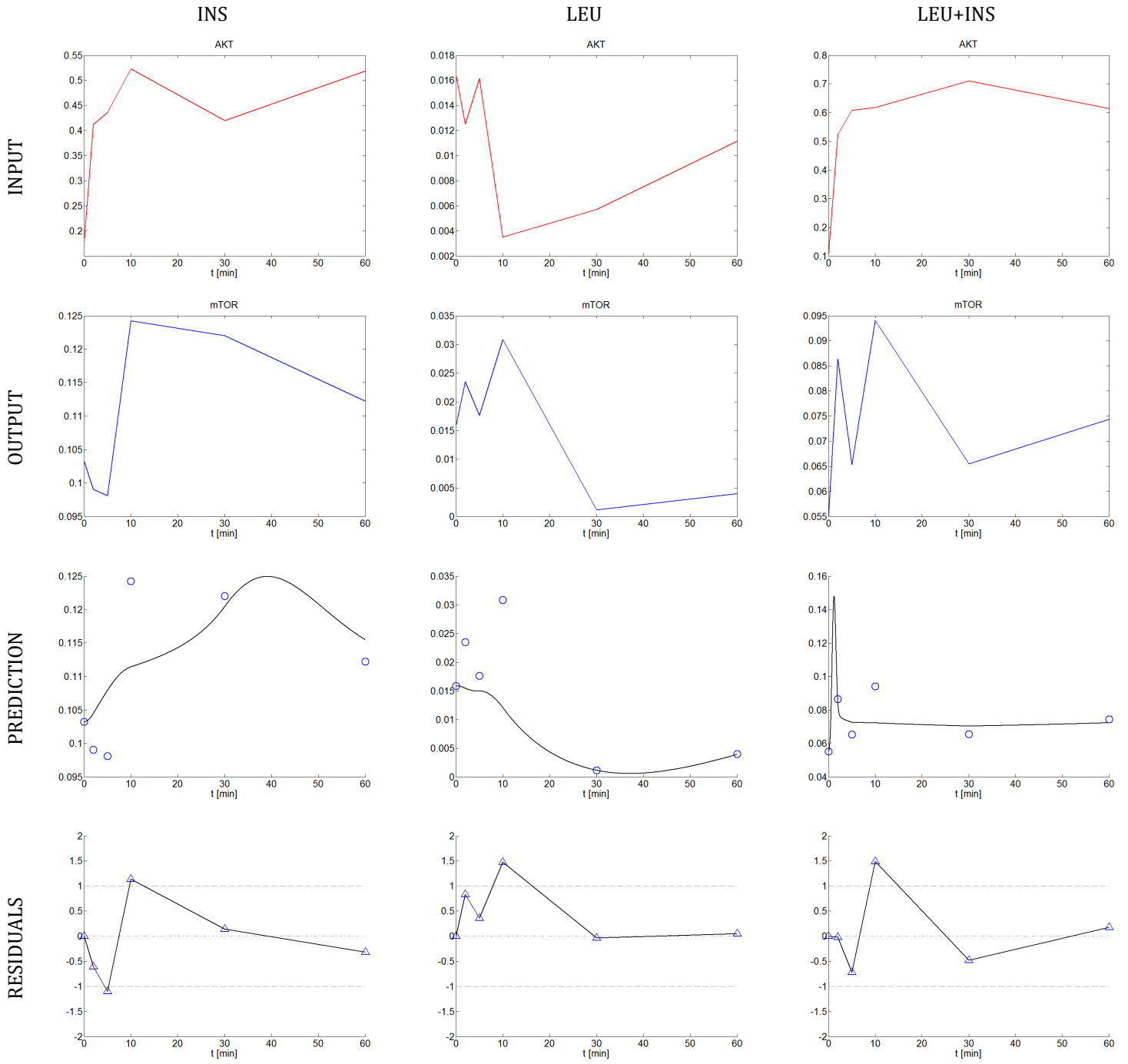
## Grafici

### sotto-sistema AKT-GSK3 $\beta$

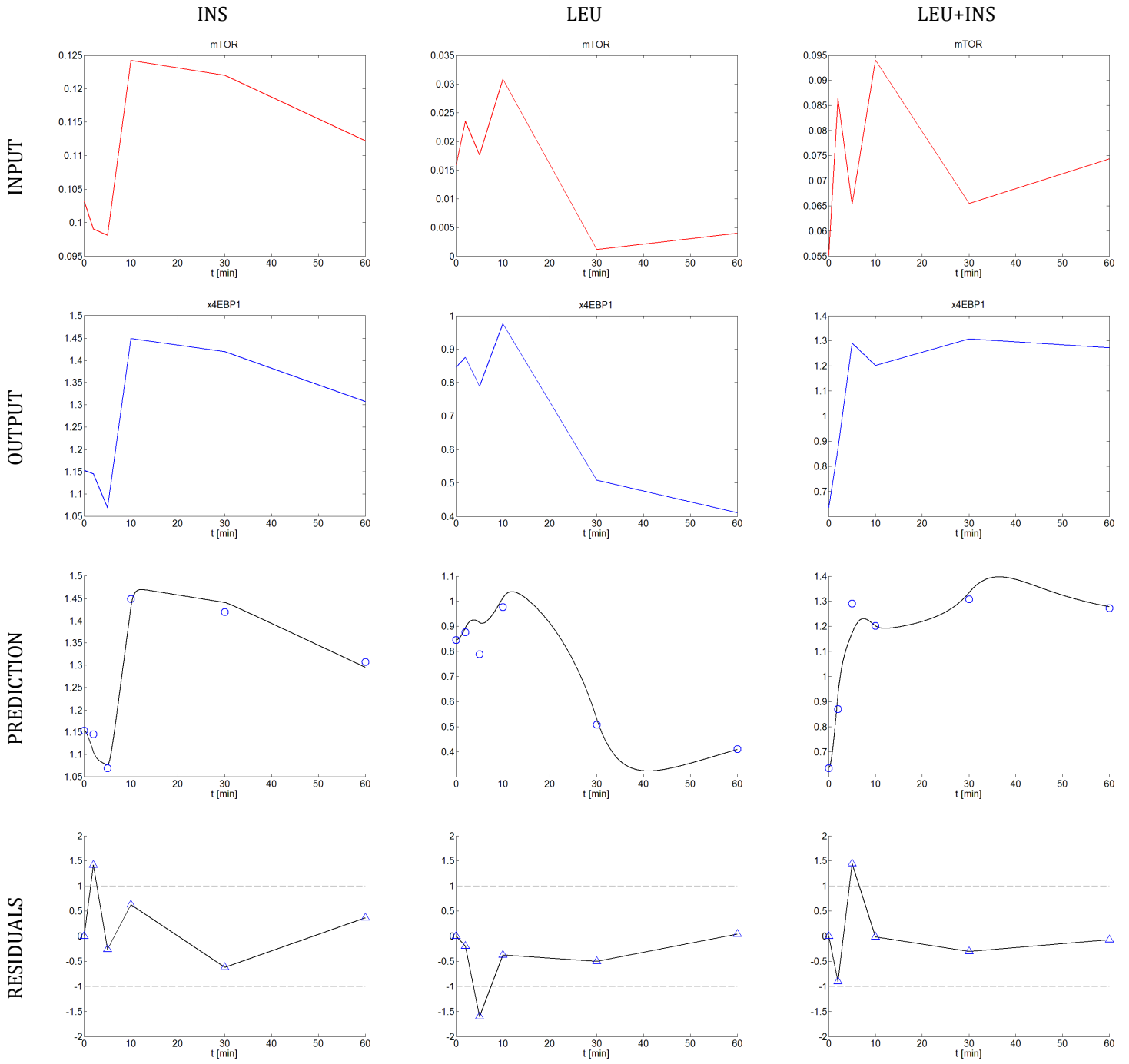




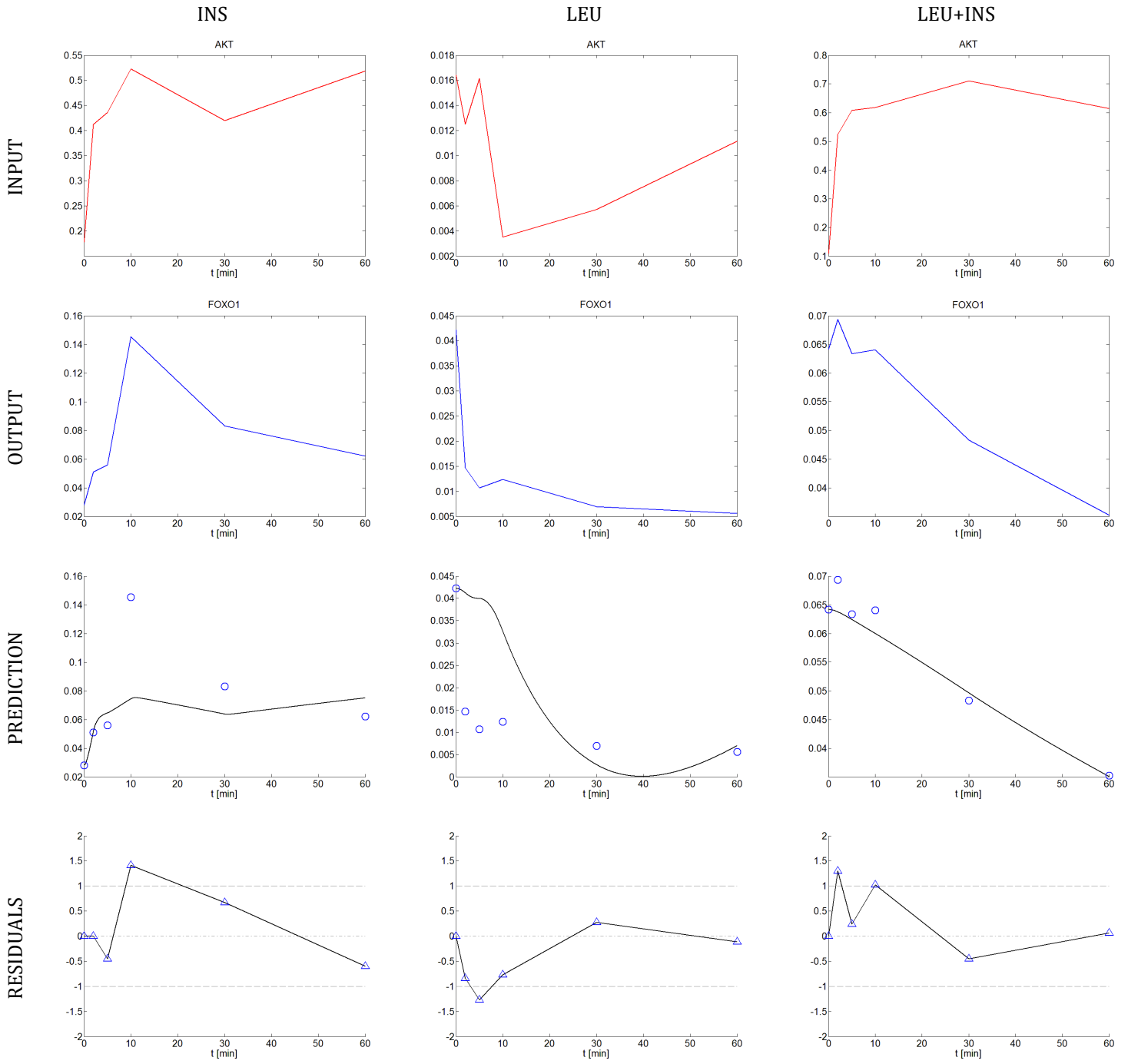
## sotto-sistema AKT-mTOR



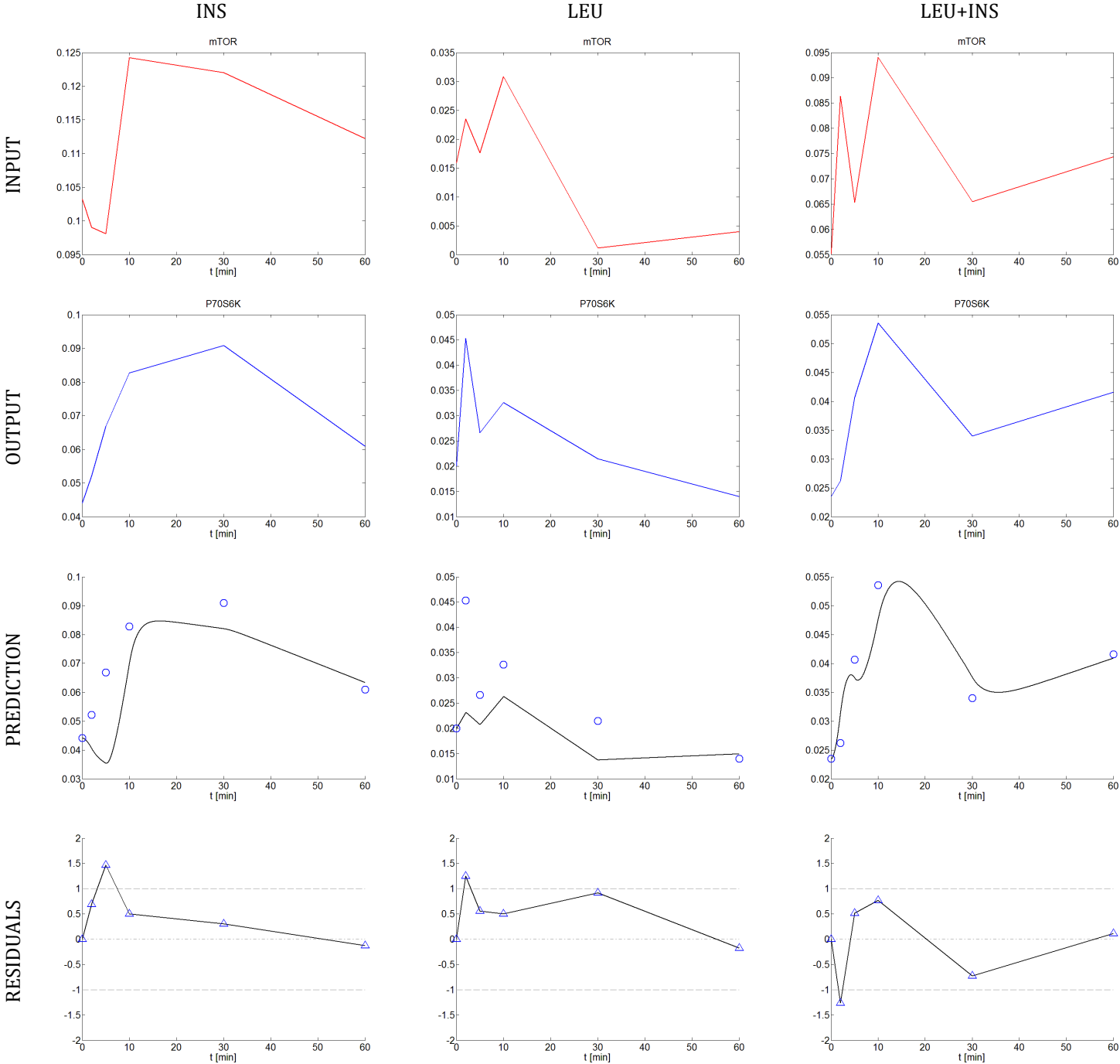
sotto-sistema mTOR-4EBP1



## sotto-sistema AKT-FOXO1



sotto-sistema mTOR-P70S6K



## Modello di misura $Y = mX$

Tabella

sotto-sistema	trattamento	F	g	h	WRSS	$\hat{\sigma}$
AKT-GSK3 $\beta$	INS	0,113 (293%)	2,576 (33%)	0,056 (13%)	0,024	0,089
	LEU	0,324 (178%)	3,504 (1796%)	0,131 (1452%)	0,250	0,289
	LEU+INS	0,000 (574468%)	0,990 (104%)	0,129 (101%)	0,059	0,140
AKT-mTOR	INS	0,000 (82801%)	0,265 (458%)	0,126 (47%)	0,027	0,094
	LEU	0,063 (616%)	0,272 (10989%)	0,400 (10753%)	0,505	0,410
	LEU+INS	1,279 (10192%)	89,967 (6%)	0,078 (70%)	0,105	0,187
mTOR-4EBP1	INS	1,454 (90%)	1,962 (742%)	14,317 (659%)	0,002	0,024
	LEU	0,189 (59%)	11,748 (60%)	1,546 (45%)	0,030	0,100
	LEU+INS	0,002 (12037%)	3,077 (172%)	3,240 (104%)	0,052	0,131
AKT-FOXO1	INS	1,085 (301%)	0,724 (1089%)	0,307 (952%)	0,360	0,347
	LEU	0,054 (1033%)	0,061 (174152%)	4,209 (173749%)	14,018	2,162
	LEU+INS	0,000 #####	0,020 (9457%)	0,000 #####	0,014	0,068
mTOR-P70S6K	INS	0,538 (917%)	0,429 (10800%)	2,799 (10526%)	0,306	0,319
	LEU	12,619 (2789%)	0,862 (91797%)	6,234 (89463%)	0,461	0,392
	LEU+INS	0,292 (426%)	0,266 (5633%)	1,138 (5537%)	0,065	0,147

I grafici corrispondenti al modello di misura  $Y = mX$  non vengono presentati perché risultano generalmente analoghi a quelli corrispondenti al modello di misura  $Y = mX + q$ . Ovviamente i risultati per il modello  $Y = mX + q$  per cui si ha  $f \geq 0$  vengono mantenuti per il modello  $Y = mX$ .

## 6.2 Pattern Expectation-Maximization

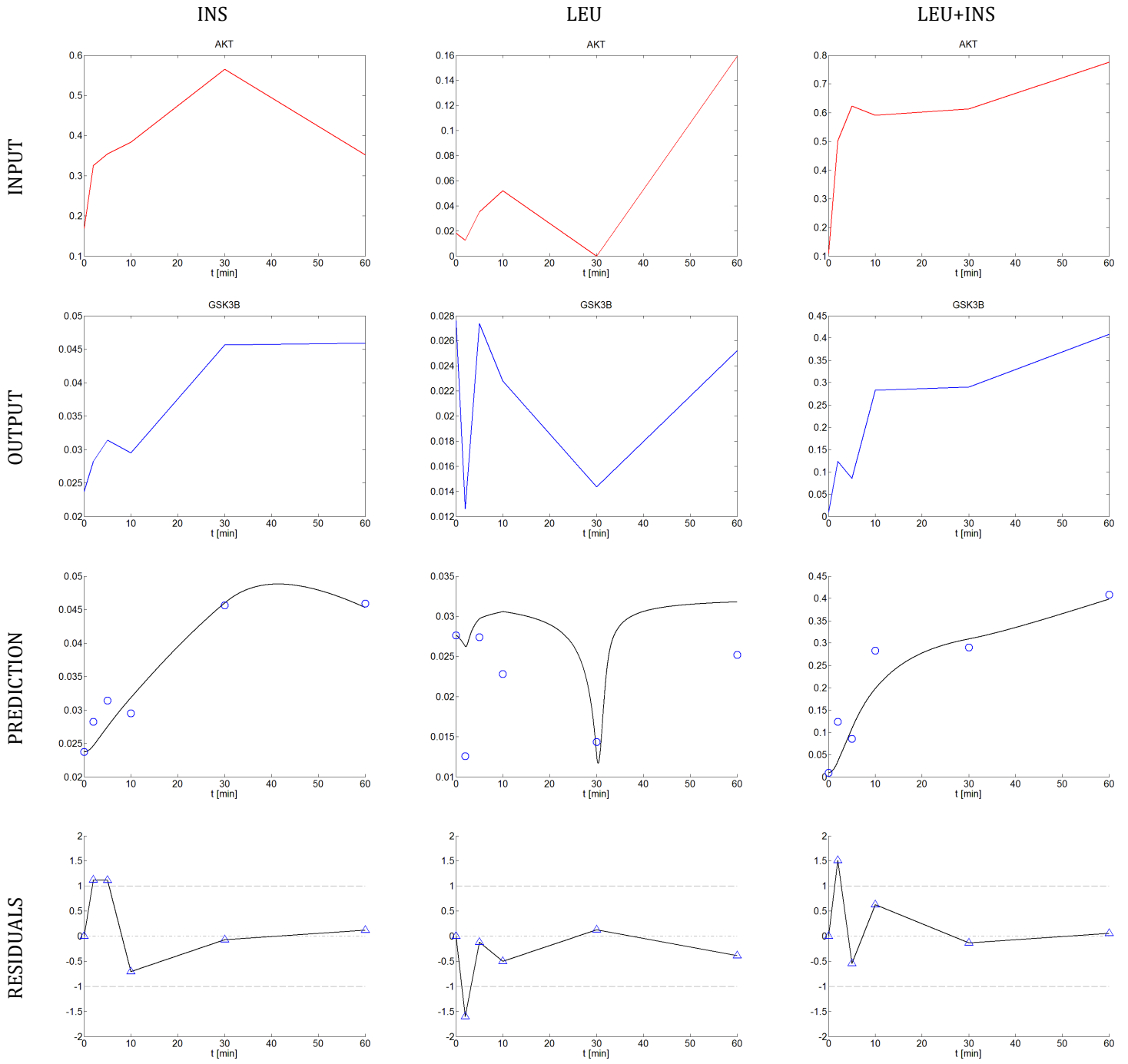
Modello di misura  $Y = mX + q$

Tabella

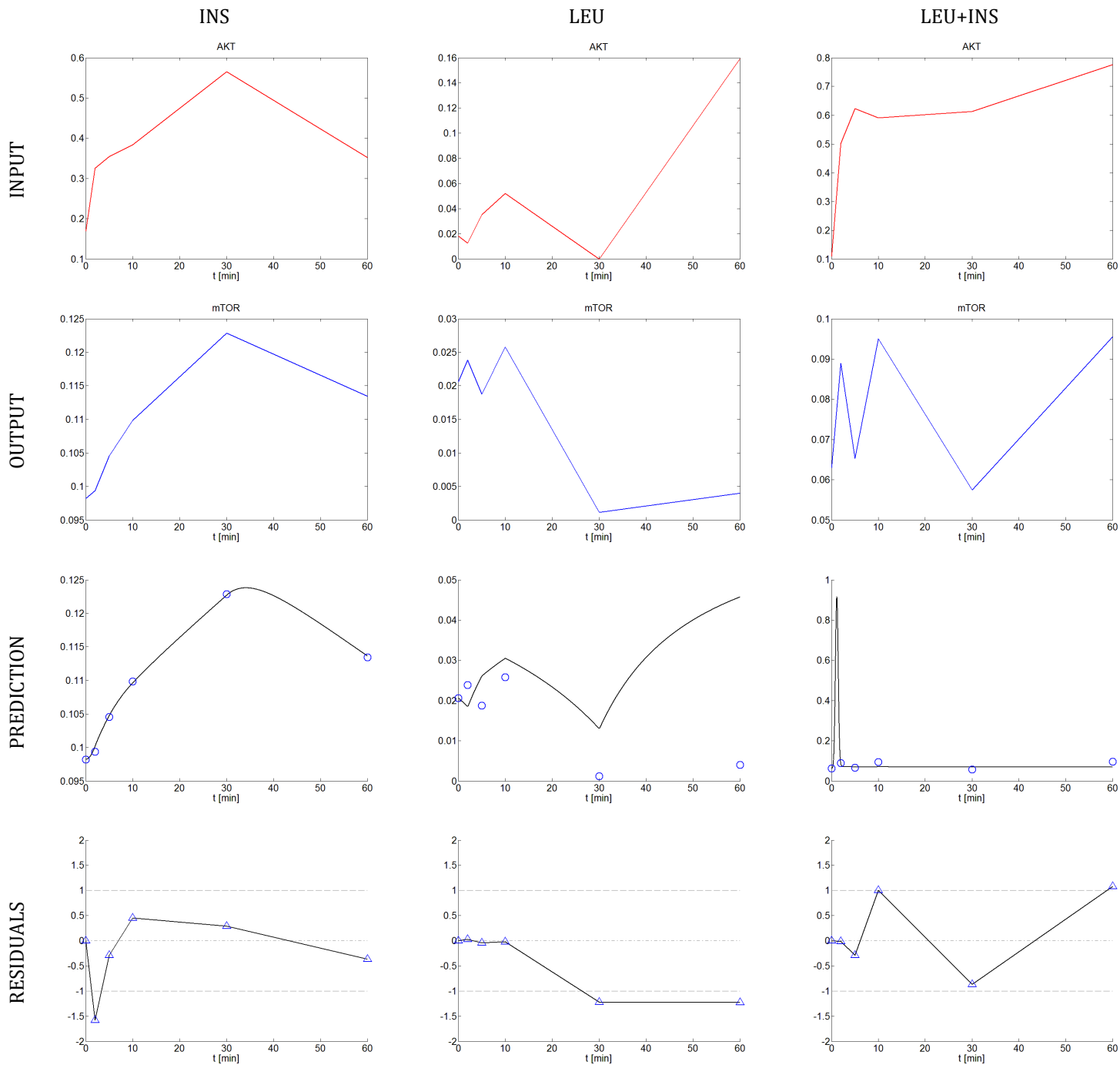
sotto-sistema	trattamento	F	g	h	WRSS	$\hat{\sigma}$
AKT-GSK3 $\beta$	INS	0,039 (492%)	0,086 (530%)	0,098 (392%)	0,038	0,112
	LEU	0,451 (492%)	160,168 (530%)	0,032 (392%)	1,374	0,677
	LEU+INS	0,106 (318%)	0,003 (15456%)	20,964 (15454%)	0,687	0,478
AKT-mTOR	INS	0,191 (75%)	0,013 (3004%)	1,180 (2753%)	0,000	0,005
	LEU	11,775 (8176%)	101,604 (3675%)	0,070 (4953%)	216,794	8,501
	LEU+INS	-19,762 (456%)	58,514 (484%)	0,067 (8%)	0,199	0,258
mTOR-4EBP1	INS	-0,535 (1043%)	12,340 (468%)	2,324 (184%)	0,021	0,084
	LEU	0,153 (135%)	0,492 (4008%)	7,877 (3685%)	0,038	0,113
	LEU+INS	-1,168 (39%)	18,118 (39%)	1,234 (9%)	0,023	0,088
AKT-FOXO1	INS	0,200 (829%)	0,005 (78991%)	12,738 (78841%)	0,578	0,439
	LEU	-0,084 (913%)	4,664 (461%)	0,004 (583%)	13,837	2,148
	LEU+INS	-0,307 (124%)	0,415 (128%)	0,064 (3%)	0,002	0,027
mTOR-P70S6K	INS	-31,957 (60%)	324,092 (62%)	0,077 (11%)	0,035	0,108
	LEU	0,180 (866%)	0,347 (23899%)	0,488 (23110%)	0,214	0,267
	LEU+INS	-0,501 (360%)	11,954 (197%)	0,067 (108%)	0,147	0,222

## Grafici

### sotto-sistema AKT-GSK3 $\beta$

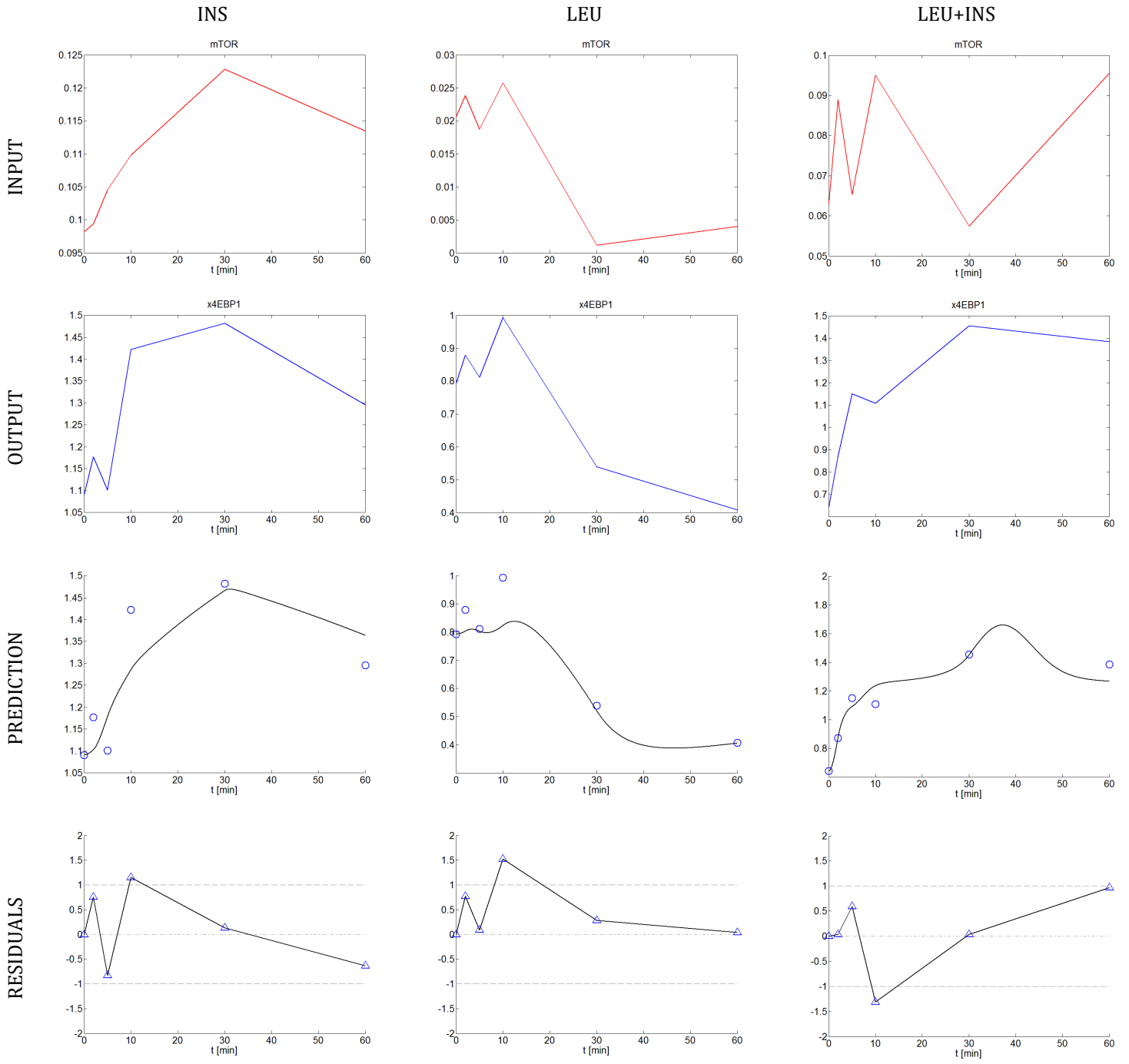


sotto-sistema AKT-mTOR

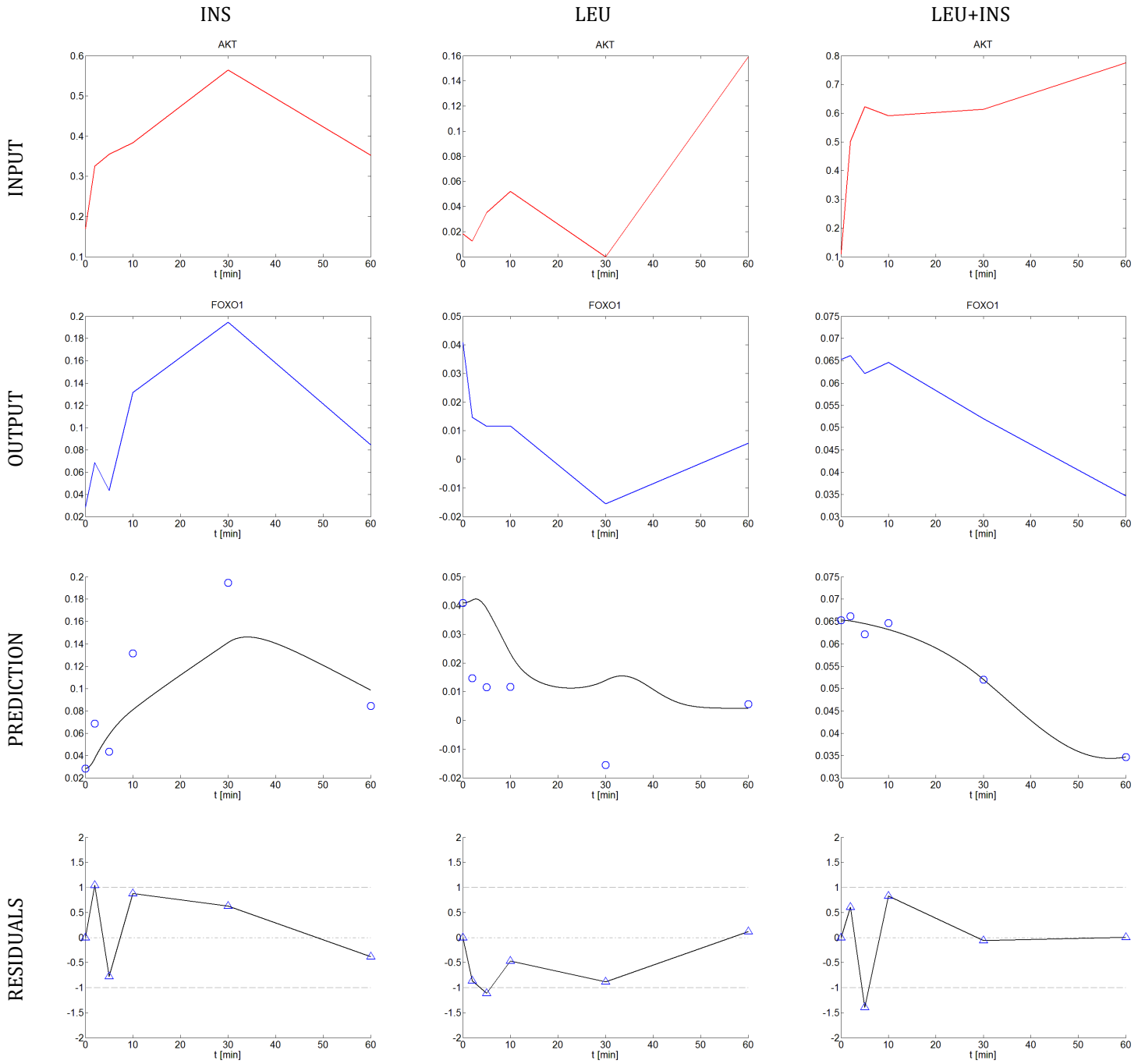




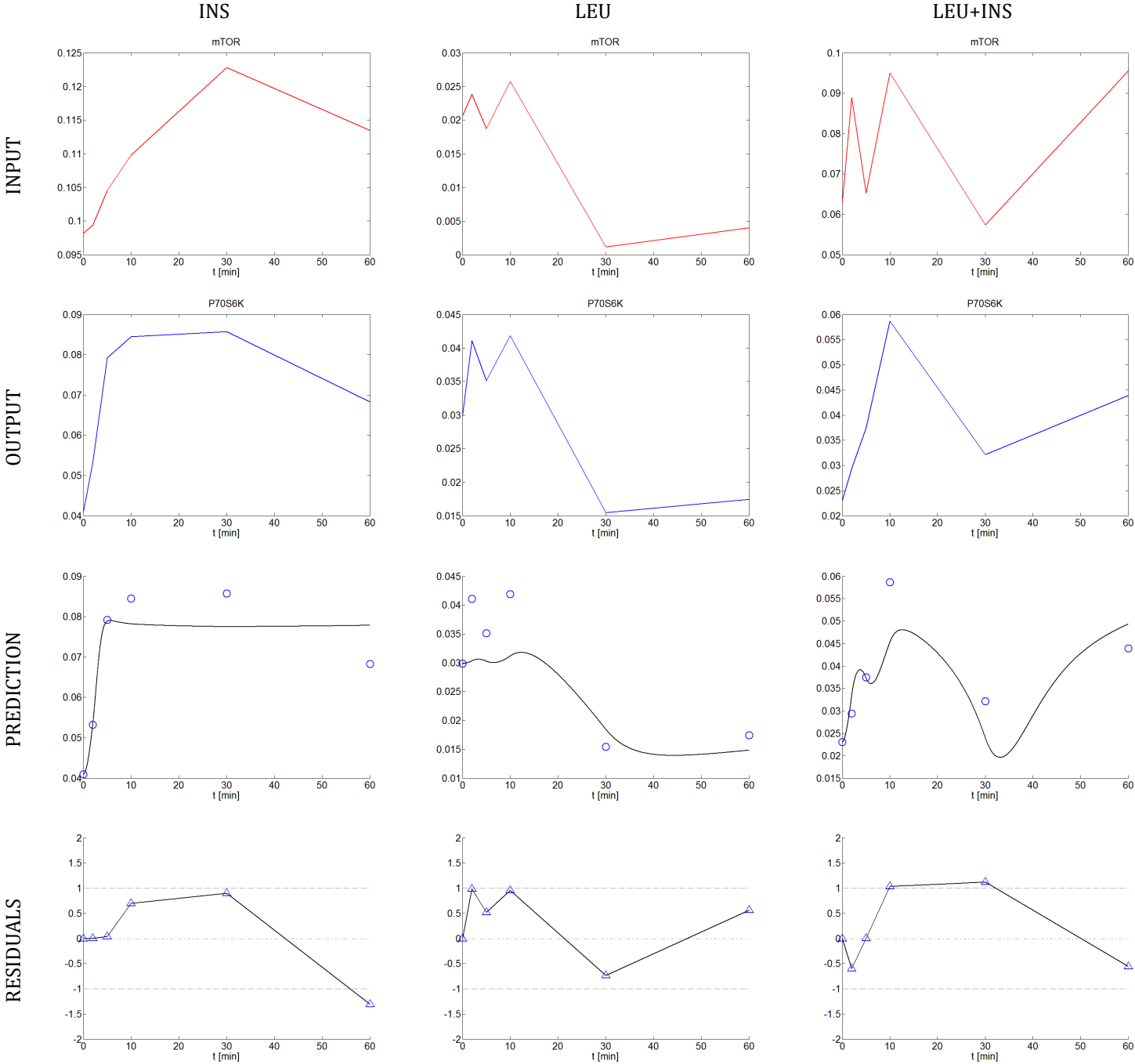
sotto-sistema mTOR-4EBP1



### sotto-sistema AKT-FOXO1



sotto-sistema mTOR-P70S6K



**Modello di misura  $Y = mX$**

Tabella

sotto-sistema	trattamento	F	g	h	WRSS	$\hat{\sigma}$
AKT-GSK3 $\beta$	INS	0,039 (492%)	0,086 (530%)	0,098 (392%)	0,038	0,112
	LEU	0,451 (492%)	160,168 (530%)	0,032 (392%)	1,374	0,677
	LEU+INS	0,106 (318%)	0,003 (15456%)	20,964 (15454%)	0,687	0,478
AKT-mTOR	INS	0,191 (75%)	0,013 (3004%)	1,180 (2753%)	0,000	0,005
	LEU	11,775 (8176%)	101,604 (3675%)	0,070 (4953%)	216,794	8,501
	LEU+INS	2,495 (22519%)	103,470 (1933%)	0,076 (168%)	0,230	0,277
mTOR-4EBP1	INS	0,698 (744%)	4,496 (1094%)	5,615 (847%)	0,021	0,084
	LEU	0,153 (135%)	0,492 (4008%)	7,877 (3685%)	0,038	0,113
	LEU+INS	0,000 #####	0,983 (224%)	4,960 (239%)	0,150	0,224
AKT-FOXO1	INS	0,200 (829%)	0,005 (78991%)	12,738 (78841%)	0,578	0,439
	LEU	0,758 (217%)	16,570 (295%)	0,000 (23221%)	31,245	3,227
	LEU+INS	0,000 #####	0,020 (1945%)	0,000 #####	0,010	0,056
mTOR-P70S6K	INS	0,614 (794%)	0,539 (8917%)	2,820 (8625%)	0,262	0,296
	LEU	0,180 (866%)	0,347 (23899%)	0,488 (23110%)	0,214	0,267
	LEU+INS	0,292 (674%)	0,266 (8024%)	1,138 (7861%)	0,198	0,257

I grafici corrispondenti al modello di misura  $Y = mX$  non vengono presentati perché risultano generalmente analoghi a quelli corrispondenti al modello di misura  $Y = mX + q$ . Ovviamente i risultati per il modello  $Y = mX + q$  per cui si ha  $f \geq 0$  vengono mantenuti per il modello  $Y = mX$ .

### 6.3 Discussione dei risultati

Le stime dei parametri ottenute, presentate in questo capitolo, sono generalmente caratterizzate da coefficienti di variazione piuttosto elevati, a dispetto di corrispondenti predizioni dell'uscita in grado di descrivere adeguatamente i dati sperimentali. Ciò è giustificato dal fatto che ogni serie temporale è costituita da un numero ridotto di campioni ( $N_C = 6$ ).

Per i dati medi, generalmente i parametri sono identificati con una precisione maggiore (cioè con coefficienti di variazione minori) di quanto avviene per i pattern Expectation-Maximization. Questa considerazione suggerisce che l'algoritmo Expectation-Maximization descritto nel paragrafo 3.3 dovrà essere opportunamente migliorato, in particolare ottimizzando il criterio di esclusione degli outliers (condizioni (\*)).

Un punto importante che resta da completare riguarda le unità di misura da assegnare ai parametri. Infatti, per indicare la quantità di una certa proteina, viene utilizzato il termine generico 'quantità', senza precisare se si tratti di  $\mu\text{M}$  o altro. Ovviamente questa indeterminazione si ripercuote nei parametri, impedendo di definirne la precisa unità di misura. La questione potrà essere risolta quando saranno note le curve di calibrazione, in grado di correlare in maniera precisa quantità e corrispondente intensità misurata. Un altro problema presente, legato al significato e alla confrontabilità dei parametri, è il fatto che i modelli compartimentali descritti nel capitolo 4 e ai quali si riferiscono appunto le stime in questo capitolo, descrivono correttamente il comportamento dei sotto-sistemi analizzati solo nell'ipotesi che le misure  $Y_1$ ,  $Y_2$  e  $Y_3$  rappresentino le quantità effettive delle corrispondenti molecole all'interno della cellula. Tuttavia, utilizzando la tecnica di misura descritta nel paragrafo 2.2, ciò che viene misurato non è l'effettiva quantità di una certa molecola all'interno della cellula ma bensì la quantità presente nel lisato caricato nel blot. Ipotizzando di poter dividere quest'ultima quantità per il volume del lisato, si potrebbe operare in termini di concentrazione, normalizzando le quantità e risolvendo in un certo senso il problema (assumendo che la concentrazione calcolata sia la stessa di quella presente all'interno della cellula). Tuttavia, come discusso nel sotto-paragrafo 2.2.5, il volume (in realtà la massa) del lisato caricato non è noto in maniera precisa, ma si conoscono solo dei valori di riferimento (corrispondente massa pari a  $10 \mu\text{g}$  oppure a  $89 \mu\text{g}$ ) da cui i valori effettivi si discostano di un errore dovuto all'applicazione manuale. Quest'ultimo problema potrà essere risolto ad esempio ottenendo direttamente misure di concentrazione anziché di quantità.

## **Capitolo 7**

### **CONCLUSIONI**

## BIBLIOGRAFIA

[1] Ahmad R. Sedaghat, Arthur Sherman and Michael J. Quon.

**A mathematical model of metabolic insulin signaling pathways.**

doi:10.1152/ajpendo.00571.2001 *Am J Physiol Endocrinol Metab* 283:E1084-E1101, 2002. First published 2 July 2002.

[2] Baston Chiara.

**Modelli matematici del signaling dell'insulina.**

(2010).

[3] Ming Wu, Xuerui Yang, Christina Chan.

**A Dynamic Analysis of IRS-PKR Signaling in Liver Cells: A Discrete Modeling Approach.**

(2009).

[4] Lippincott Williams & Wilkins.

**The Insulin Signaling Network and Insulin Action.**

Diabetes Mellitus: A Fundamental and Clinical Text 3rd Edition (2004).